DEX-0279 BG

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number
WO 02/40535 A2

(51) International Patent Classification[7]: C07K 14/47, C12N 15/12, C12Q 1/68, C07K 16/18, G01N 33/50

(21) International Application Number: PCT/US01/45011

(22) International Filing Date: 20 November 2001 (20.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/252,061 20 November 2000 (20.11.2000) US
60/253,257 27 November 2000 (27.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/252,061 (CIP)
Filed on 20 November 2000 (20.11.2000)
US 60/253,257 (CIP)
Filed on 27 November 2000 (27.11.2000)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SALCEDA, Susana** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **CAFFERKEY, Robert** [IE/US]; 350 Elan Village Lane, Apartment 218, San Jose, CA 95134 (US). **SUN, Yongming** [US/US]; 869 S. Winchester Boulevard, Apartment 260, San Jose, CA 95128 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way, #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/40535 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovary tissue, identifying ovary tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application Serial No. 60/252,061 filed November 20, 2000, and U.S. Provisional Application Serial No. 60/253,257 filed November 27, 2000, which are herein incorporated by reference in their entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovary tissue, identifying ovary tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

BACKGROUND OF THE INVENTION

Cancer of the ovaries is the fourth-most cause of cancer death in women in the United States, with more than 23,000 new cases and roughly 14,000 deaths predicted for the year 2001. Shridhar, V. et al., Cancer Res. 61(15): 5895-904 (2001); Memarzadeh, S. & Berek, J. S., J. Reprod. Med. 46(7): 621-29 (2001). The incidence of ovarian cancer is of serious concern worldwide, with an estimated 191,000 new cases predicted anually. Runnebaum, I. B. & Stickeler, E., J. Cancer Res. Clin. Oncol. 127(2): 73-79 (2001). Because women with ovarian cancer are typically asypmtomatic until the disease has metastasized, and because effective screening for ovarian

cancer is not available, roughly 70% of women present with an advanced stage of the cancer, with a five-year survival rate of ~25-30% at that stage. Memarzadeh, S. & Berek, J. S., supra; Nunns, D. et al., Obstet. Gynecol. Surv. 55(12): 746-51. Conversely, women diagnosed with early stage ovarian cancer enjoy considerably higher survival rates. Werness, B. A. & Eltabbakh, G. H., Int'l. J. Gynecol. Pathol. 20(1): 48-63 (2001).

Although our understanding of the etiology of ovarian cancer is incomplete, the results of extensive research in this area point to a combination of age, genetics, reproductive, and dietary/environmental factors. Age is a key risk factor in the development of ovarian cancer: while the risk for developing ovarian cancer before the age of 30 is slim, the incidence of ovarian cancer rises linearly between ages 30 to 50, increasing at a slower rate thereafter, with the highest incidence being among septagenarian women. Jeanne M. Schilder et al., Heriditary Ovarian Cancer: Clinical Syndromes and Management, in Ovarian Cancer 182 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001).

With respect to genetic factors, a family history of ovarian cancer is the most significant risk factor in the development of the disease, with that risk depending on the number of affected family members, the degree of their relationship to the woman, and which particular first degree relatives are affected by the disease. Id. Mutations in several genes have been associated with ovarian cancer, including BRCA1 and BRCA2, both of which play a key role in the development of breast cancer, as well as hMSH2 and hMLH1, both of which are associated with heriditary non-polyposis ovary cancer. Katherine Y. Look, Epidemiology, Etiology, and Screening of Ovarian Cancer, in Ovarian Cancer 169, 171-73 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). BRCA1, located on chromosome 17, and BRCA2, located on chromosome 13, are tumor supressor genes implicated in DNA repair; mutations in these genes are linked to roughly 10% of ovarian cancers. Id. at 171-72; Schilder et al., supra at 185-86. hMSH2 and hMLH1 are associated with DNA mismatch repair, and are located on chromsomes 2 and 3, respectively; it has been reported that roughly 3% of heriditary ovarian carcinomas are due to

mutations in these genes. Look, supra at 173; Schilder et al., supra at 184, 188-89.

Reproductive factors have also been associated with an increased or reduced risk of ovarian cancer. Late menopause, nulliparity, and early age at menarche have all been linked with an elevated risk of ovarian cancer. Schilder et al., supra at 182. One theory hypothesizes that these factors increase the number of ovulatory cycles over the course of a woman's life, leading to "incessant ovulation," which is thought to be the primary cause of mutations to the ovarian epithelium. Id.; Laura J. Havrilesky & Andrew Berchuck, Molecular Alterations in Sporadic Ovarian Cancer, in Ovarian Cancer 25 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). The mutations may be explained by the fact that ovulation results in the destruction and repair of that epithelium, necessitating increased cell division, thereby increasing the possibility that an undesried mutation will occur. Id. Support for this theory may be found in the fact pregnancy, lactation, and the use of oral contraceptives, all of which suppress ovulation, confer a protective effect with respect to developing ovarian cancer. Id.

Among dietary/environmental factors, there would appear to be an association between high intake of animal fat or red meat and ovarian cancer, while the antioxidant Vitamin A, which prevents free radical formation and also assists in maintaining normal cellular differentiation, may offer a protective effect. Look, supra at 169. Reports have also associated asbestos and hydrous magnesium trisilicate (talc), the latter of which may be present in diaphragms and sanitary napkins. Id. at 169-70.

Current screening procedures for ovarian cancer, while of some utility, are quite limited in their diagnostic ability, a problem that is particularly acute at early stages of cancer progression when the disease is typically asymptomatic yet is most readily treated. Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 166 (1998); Memarzadeh & Berek, supra; Runnebaum & Stickeler, supra; Werness & Eltabbakh, supra. Commonly used screening tests include bimanual rectovaginal pelvic examination,

radioimmunoassay to detect the CA-125 serum tumor marker, and transvaginal ultrasonography. Burdette, supra at 166.

Pelvic examination has failed to yield adequate numbers of early diagnoses, and the other methods are not sufficiently accurate. Id. One study reported that only 15% of patients who suffered from ovarian cancer were diagnosed with the disease at the time of their pelvic examination. Look, supra at 174. Moreover, the CA-125 test is prone to giving false positives in pre-menopausal women and has been reported to be of low predictive value in post-menopausal women. Id. at 174-75. Although transvaginal ultrasonographyis now the preferred procedure for screening for ovarian cancer, it is unable to distinguish reliably between benign and malignant tumors, and also cannot locate primary peritoneal malignancies or ovarian cancer if the ovary size is normal. Schilder et al., supra at 194-95. While genetic testing for mutations of the BRCA1, BRCA2, hMSH2, and hMLH1 genes is now available, these tests may be too costly for some patients and may also yield false negative or indeterminate results. Schilder et al., supra at 191-94.

The staging of ovarian cancer, which is accomplished through surgical exploration, is crucial in determining the course of treatment and management of the disease. AJCC Cancer Staging Handbook 187 (Irvin D. Fleming et al. eds., 5th ed. 1998); Burdette, supra at 170; Memarzadeh & Berek, supra; Shridhar et al., supra. Staging is performed by reference to the classification system developed by the International Federation of Gynecology and Obstetrics. David H. Moore, Primary Surgical Management of Early Epithelial Ovarian Carcinoma, in Ovarian Cancer 203 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001); Fleming et al. eds., supra at 188. Stage I ovarian cancer is characterized by tumor growth that is limited to the ovaries and is comprised of three substages. Id. In substage IA, tumor growth is limited to one ovary, there is no tumor on the external surface of the ovary, the ovarian capsule is intact, and no malignant cells are present in ascites or peritoneal washings. Id. Substage IB is identical to A1, except that tumor growth is limited to both ovaries. Id. Substage IC

refers to the presence of tumor growth limited to one or both ovaries, and also includes one or more of the following characteristics: capsule rupture, tumor growth on the surface of one or both ovaries, and malignant cells present in ascites or peritoneal washings. Id.

Stage II ovarian cancer refers to tumor growth involving one or both ovaries, along with pelvic extension. Id. Substage IIA involves extension and/or implants on the uterus and/or fallopian tubes, with no malignant cells in the ascites or peritoneal washings, while substage IIB involves extension into other pelvic organs and tissues, again with no malignant cells in the ascites or peritoneal washings. Id. Substage IIC involves pelvic extension as in IIA or IIB, but with malignant cells in the ascites or peritoneal washings. Id.

Stage III ovarian cancer involves tumor growth in one or both ovaries, with peritoneal metastasis beyond the pelvis confirmed by microscope and/or metastasis in the regional lymph nodes. Id. Substage IIIA is characterized by microscopic peritoneal metastasis outside the pelvis, with substage IIIB involving macroscopic peritoneal metastasis outside the pelvis 2 cm or less in greatest dimension. Id. Substage IIIC is identical to IIIB, except that the metastisis is greater than 2 cm in greatest dimesion and may include regional lymph node metastasis. Id. Lastly, Stage IV refers to the presence distant metastasis, excluding peritoneal metastasis. Id.

While surgical staging is currently the benchmark for assessing the management and treatment of ovarian cancer, it suffers from considerable drawbacks, including the invasiveness of the procedure, the potential for complications, as well as the potential for inaccuracy. Moore, supra at 206-208, 213. In view of these limitations, attention has turned to developing alternative staging methodologies through understanding differential gene expression in various stages of ovarian cancer and by obtaining various biomarkers to help better assess the progression of the disease. Vartiainen, J. et al., Int'l J. Cancer, 95(5): 313-16 (2001); Shridhar et al. supra; Baekelandt, M. et al., J. Clin. Oncol. 18(22): 3775-81.

The treatment of ovarian cancer typically involves a multiprong attack, with surgical intervention serving as the foundation of treatment. Dennis S. Chi & William J. Hoskins, Primary Surgical Management of Advanced Epithelial Ovarian Cancer, in Ovarian Cancer 241 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). For example, in the case of epithelial ovarian cancer, which accounts for ~90% of cases of ovarian cancer, treatment typically consists of: (1) cytoreductive surgery, including total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy, and lymphadenectomy, followed by (2) adjuvant chemotherapy with paclitaxel and either cisplatin or carboplatin. Eltabbakh, G.H. & Awtrey, C.S., Expert Op. Pharmacother. 2(10): 109-24. Despite a clinical response rate of 80% to the adjuvant therapy, most patients experience tumor recurrence within three years of treatment. Id. Certain patients may undergo a second cytoreductive surgery and/or second-line chemotherapy. Memarzadeh & Berek, supra.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of ovarian cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop ovarian cancer, for diagnosing ovarian cancer, for monitoring the progression of the disease, for staging the ovarian cancer, for determining whether the ovarian cancer has metastasized, and for imaging the ovarian cancer. There is also a need for better treatment of ovarian cancer.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovaries; identify and monitor ovary tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered ovary tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to ovary cells and/or ovary tissue. These ovary specific nucleic acids (OSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the OSNA is genomic DNA, then the OSNA is an ovary specific gene (OSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to ovary. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 94 through 167. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 93. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding an OSP, or that selectively hybridize or exhibit substantial sequence similarity to an OSNA, as well as allelic variants of a nucleic acid molecule encoding an OSP, and allelic variants of an OSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes an OSP or that comprises a part of a nucleic acid sequence of an OSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of an OSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of an OSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic

acid molecule encodes all or a fragment of an OSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of an OSNA.

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is an OSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of an OSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovaries. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring ovary tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered ovary tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovaries. The invention provides methods of using the polypeptides of the invention to identify and/or monitor ovary tissue, and to produce engineered ovary tissue.

The agonists and antagonists of the instant invention may be used to treat ovarian cancer and non-cancerous disease states in ovaries and to produce engineered ovary tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in

the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may

comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid

molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are

not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence.

In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) – 0.63 (% formamide) - (600/l)

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping

the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:

$T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction } G+C) - (600/N)$,

wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAS. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity

with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al., Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding an OSP or is an OSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al., Technique* 1: 11-15 (1989) and Caldwell *et al., PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al., Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR

reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression

vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises an OSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally

associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial

identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$--, and –CH_2SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid

sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino

acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, -N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted

upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.,* Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)

Cost to extend a gap: 1 (default
Max. alignments: 100 (default)
Word size: 11 (default)
No. of descriptions: 100 (default)
Penalty Matrix: BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an $F(ab')_2$ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single

arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.,* Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.,* Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.*, Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that

purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "ovary specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the ovary as compared to other tissues in the body. In a preferred embodiment, a "ovary specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "ovary specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the ovary or to ovary cells or tissue or that are derived from such nucleic acid molecules. These isolated ovary specific nucleic acids (OSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to ovary, an ovary-specific polypeptide (OSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 94 through 167. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 93.

AN OSNA may be derived from a human or from another animal. In a preferred embodiment, the OSNA is derived from a human or other mammal. In a more preferred embodiment, the OSNA is derived from a human or other primate. In an even more preferred embodiment, the OSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding an OSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode an OSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes an OSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 94 through 167. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 93.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under moderate stringency conditions. In a more preferred

embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 94 through 167. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 93. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding an OSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human OSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 94 through 167. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding an OSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 94 through 167, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding an OSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding an OSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to an OSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 93. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity

with an OSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 93, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with an OSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with an OSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to an OSNA or to a nucleic acid molecule encoding an OSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the OSNA or the nucleic acid molecule encoding an OSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 94 through 167 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 93. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the OSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is

experimentally produced by directed mutation of an OSNA. Further, the substantially similar nucleic acid molecule may or may not be an OSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is an OSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of an OSNA or a nucleic acid encoding an OSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes an OSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is an OSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 93. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is an OSP. However, in a preferred embodiment, the part encodes an OSP. In one aspect, the invention comprises a part of an OSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to an OSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of an OSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes an OSP. A part comprises at least 10 nucleotides, more preferably at least 15,

17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as -^{32}P-dATP, -^{32}P-dCTP, -^{32}P-dGTP, -^{32}P-dTTP, -^{32}P-3'dATP, -^{32}P-ATP, -^{32}P-CTP, -^{32}P-GTP, -^{32}P-UTP, -^{35}S-dATP, α-^{35}S-GTP, α-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3'

hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.*, Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.),

Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See* Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S.

Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases

do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably,

detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of an OSNA, such as deletions, insertions, translocations, and duplications of the OSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify OSNA in, and isolate OSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^+-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to OSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding an OSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 94 through 167. In another preferred embodiment, the probe or primer is derived from an OSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 93.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular

Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See*, *e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See*, *e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See*, *e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention

can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides

encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their

derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda* , *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-

transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of

the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast _-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early

gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the OSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PltetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to

be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from

those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,*. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide OSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database http://www.abrf.org/ABRF/Research Committees/deltamass/deltamass.html (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29: 332-335 (2001) and http://www.glycosuite.com/ (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999) and http://www.cbs.dtu.dk/databases/OGLYCBASE/ (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) and http://www.cbs.dtu.dk/ databases/PhosphoBase/ (accessed October 19, 2001); or http://pir.georgetown.edu/ pirwww/search/textresid.html (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from ovary are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human ovary cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli,* in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^{+} or K^{+}, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols

(BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/ New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/ New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al., Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al., Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is an ovary specific polypeptide (OSP). In an even more preferred embodiment, the

polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 94 through 167. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of an OSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 94 through 167. A polypeptide that comprises only a fragment of an entire OSP may or may not be a polypeptide that is also an OSP. For instance, a full-length polypeptide may be ovary-specific, while a fragment thereof may be found in other tissues as well as in ovary. A polypeptide that is not an OSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-OSP antibodies. However, in a preferred embodiment, the part or fragment is an OSP. Methods of determining whether a polypeptide is an OSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for

the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, an OSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See*, *e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably an OSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably an OSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be ovary-specific. In a preferred embodiment, the mutein is ovary-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 94 through 167. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is ovary-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo*

mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to an OSP. In an even more preferred embodiment, the polypeptide is homologous to an OSP selected from the group having an amino acid sequence of SEQ ID NO: 94 through 167. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to an OSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 94 through 167. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to an OSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to an OSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the OSNA is selected from the group consisting of SEQ ID NO: 1 through 93. In another preferred

embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes an OSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the OSP is selected from the group consisting of SEQ ID NO: 94 through 167.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 94 through 167. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the OSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of an OSP. Further, the homologous protein may or may not encode polypeptide that is an OSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is an OSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the

binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding an OSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 94 through 167. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 93.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an OSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 94 through 167, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983);

Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591,

BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-OSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999),

incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an OSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 94 through 167. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to an OSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$-- and –CH_2SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of an OSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.*, Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is an OSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 94 through 167, or is a mutein, homologous

polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 93, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 93.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See*, *e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation

of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbrizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in Drosophila with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by

reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast _ mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the OSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize OSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly OSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser

scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of OSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of OSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.,* Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is an OSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 94 through 167, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on an OSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of an OSP may be indicative of cancer. Differential degradation of the C or N-terminus of an OSP may also be a marker or target for anticancer therapy. For example, an OSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-OSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human ovary.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the

present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to $1 \mathrm{X} 10^{-13}$ M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are

hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al., Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al., J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular

advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*,

4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.*, Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,

Biotechnol. Appl. Biochem. 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); Limonta et al., *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'$_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated

nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions

including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate

reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al., Methods Enzymol.* 133: 331-53 (1986); Kricka *et al., J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al., J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I.

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding an OSP. In a preferred embodiment, the OSP comprises an amino

acid sequence selected from SEQ ID NO: 94 through 167, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise an OSNA of the invention, preferably an OSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 93, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human OSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.*, Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.*, Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.*, Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.*, Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.*, Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.*, a

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce ovaryies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.*, Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.*, Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.*, Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC

compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 93 and SEQ ID NO: 94 through 167 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Ovarian Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of an OSNA or an OSP in a human patient that has or may have ovarian cancer, or who is at risk of developing ovarian cancer, with the expression of an OSNA or an OSP in a normal human control. For purposes of the present invention, "expression of an OSNA" or "OSNA expression" means the quantity of OSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of an OSP" or "OSP expression" means the amount of OSP that can be measured by any method known in the art or the level of translation of an OSG OSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing ovarian cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of OSNA or OSP in cells, tissues, organs or bodily fluids compared with levels of OSNA or OSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of an OSNA or OSP in the patient versus the normal human control is associated with the presence of ovarian cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in the structure of the mRNA of an OSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in an OSP compared to an OSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the OSP or subcellular OSP localization.

In a preferred embodiment, the expression of an OSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 94 through 167, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the OSNA expression that is measured is the level of expression of an OSNA mRNA selected from SEQ ID NO: 1 through 93, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. OSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by

Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. OSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of an OSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, OSNA expression may be compared to a known control, such as normal ovary nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of an OSP is measured by determining the level of an OSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 94 through 167, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of OSNA or OSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of ovarian cancer. The expression level of an OSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the OSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the OSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to an OSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-OSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the OSP will bind to the anti-OSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-OSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the OSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an OSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure OSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-OSP antibody is attached to a solid support and an allocated amount of a labeled OSP and a sample of interest are incubated with the solid support. The amount of labeled OSP detected which is attached to the solid support can be correlated to the quantity of an OSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of an OSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more OSNAs of interest. In this approach, all or a portion of one or more OSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of OSNA or OSP includes, without limitation, ovary tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, ovary cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary ovarian cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and breast. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in OSNAs or OSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of an OSNA or OSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other OSNA or OSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular OSNA or OSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a patient suspected of having ovarian cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP and then ascertaining whether the patient has ovarian cancer from the expression level of the OSNA or OSP. In general, if high expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether ovarian cancer has metastasized in a patient. One may identify whether the ovarian cancer has metastasized by measuring the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a variety of tissues. The presence of an OSNA or OSP in a

certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of an OSNA or OSP is associated with ovarian cancer. Similarly, the presence of an OSNA or OSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of an OSNA or OSP is associated with ovarian cancer. Further, the presence of a structurally altered OSNA or OSP that is associated with ovarian cancer is also indicative of metastasis.

In general, if high expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The OSNA or OSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with ovarian cancers or other ovary related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of ovarian disorders.

Staging

The invention also provides a method of staging ovarian cancer in a human patient. The method comprises identifying a human patient having ovarian cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more OSNAs or OSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more OSNAs or OSPs is determined for each stage to obtain a standard expression level for each OSNA and OSP. Then, the OSNA or OSP expression levels are determined in a biological sample from a patient whose stage of

cancer is not known. The OSNA or OSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the OSNAs and OSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of an OSNA or OSP to determine the stage of an ovarian cancer.

Monitoring

Further provided is a method of monitoring ovarian cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the ovarian cancer. The method comprises identifying a human patient that one wants to monitor for ovarian cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more OSNAs or OSPs, and comparing the OSNA or OSP levels over time to those OSNA or OSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in an OSNA or OSP that are associated with ovarian cancer.

If increased expression of an OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of an OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of OSNAs or OSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of ovarian cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of an OSNA and/or OSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more OSNAs and/or OSPs are detected. The presence of higher (or lower) OSNA or OSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly ovarian cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more OSNAs and/or OSPs of the invention can also be monitored by analyzing levels of expression of the OSNAs and/or OSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in an OSG, thereby determining if a human with the genetic lesion is susceptible to developing ovarian cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing ovarian cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the OSGs of this invention, a chromosomal rearrangement of OSG, an aberrant modification of OSG (such as of the methylation pattern of the genomic DNA), or allelic loss of an OSG. Methods to detect such lesions in the OSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Ovarian Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a patient suspected of having or known to have a noncancerous ovarian disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP, comparing the expression level or structural alteration of the OSNA or OSP to a normal

ovary control, and then ascertaining whether the patient has a noncancerous ovarian disease. In general, if high expression relative to a control of an OSNA or OSP is indicative of a particular noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of a noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether an OSNA and/or OSP is associated with a particular noncancerous ovarian disease by obtaining ovary tissue from a patient having a noncancerous ovarian disease of interest and determining which OSNAs and/or OSPs are expressed in the tissue at either a higher or a lower level than in normal ovary tissue. In another embodiment, one may determine whether an OSNA or OSP exhibits structural alterations in a particular noncancerous ovarian disease state by obtaining ovary tissue from a patient having a noncancerous ovarian disease of interest and determining the structural alterations in one or more OSNAs and/or OSPs relative to normal ovary tissue.

Methods for Identifying Ovary Tissue

In another aspect, the invention provides methods for identifying ovary tissue. These methods are particularly useful in, *e.g.*, forensic science, ovary cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is ovary tissue or has ovary tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising ovary tissue or having ovary tissue-like characteristics, determining whether the sample expresses one or more OSNAs and/or OSPs, and, if the sample expresses one or more OSNAs and/or OSPs, concluding

that the sample comprises ovary tissue. In a preferred embodiment, the OSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 94 through 167, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 93, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses an OSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether an OSP is expressed. Determining whether a sample expresses an OSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the OSP has an amino acid sequence selected from SEQ ID NO: 94 through 167, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two OSNAs and/or OSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five OSNAs and/or OSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is ovary tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into ovary tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new ovary tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

<u>Methods for Producing and Modifying Ovary Tissue</u>

In another aspect, the invention provides methods for producing engineered ovary tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing an OSNA or an OSG into the cells, and growing the cells under conditions in

which they exhibit one or more properties of ovary tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal ovary tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered ovary tissue or cells comprises one of these cell types. In another embodiment, the engineered ovary tissue or cells comprises more than one ovary cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the ovary cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more OSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode OSPs having amino acid sequences selected from SEQ ID NO: 94 through 167, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 93, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, an OSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial ovary tissue may be used to treat patients who have lost some or all of their ovary function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises an OSNA or part thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 93, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises an OSP or fragment thereof. In a more preferred embodiment, the OSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 94 through 167, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of

the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-OSP antibody, preferably an antibody that specifically binds to an OSP having an amino acid that is selected from the group consisting of SEQ ID NO: 94 through 167, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot

injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts

tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example OSP polypeptide, fusion protein, or fragments thereof, antibodies specific for OSP, agonists, antagonists or inhibitors of OSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age,

weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention , *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of ovary function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication

incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of an OSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of an OSP are administered, for example, to complement a deficiency in the native OSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra.* In a preferred embodiment, the nucleic acid molecule encodes an OSP having the amino acid sequence of SEQ ID NO: 94 through 167, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express an OSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in OSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode an OSP having the amino acid sequence of SEQ ID NO: 94 through 167, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of an OSG antisense nucleic acid, are administered to downregulate transcription and/or translation of an OSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of an OSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to OSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the OSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding an OSP, preferably an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 93, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an OSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant OSP defect.

Protein compositions are administered, for example, to complement a deficiency in native OSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to OSP. The immune

response can be used to modulate activity of OSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate OSP.

In a preferred embodiment, the polypeptide is an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 83, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of OSP, or to target therapeutic agents to sites of OSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 93, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to an OSP or have a modulatory effect on the expression or activity of an OSP. Modulators which decrease the expression or activity of OSP (antagonists) are believed to be useful in treating ovarian cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of an OSP can also be designed, synthesized and tested for use in the imaging and treatment of ovarian cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the OSPs identified herein. Molecules identified in the library as being capable of binding to an OSP are key candidates for

further evaluation for use in the treatment of ovarian cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of an OSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of OSP is administered. Antagonists of OSP can be produced using methods generally known in the art. In particular, purified OSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of an OSP.

In other embodiments a pharmaceutical composition comprising an agonist of an OSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP comprising an amino acid sequence of SEQ ID NO: 94 through 167, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 93, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Ovary Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the ovary or to specific cells in the ovary. In a preferred embodiment, an anti-OSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if ovary tissue needs to be selectively destroyed. This would be useful for targeting and killing ovarian cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting ovary cell function.

In another embodiment, an anti-OSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring ovary function, identifying ovarian cancer tumors, and identifying noncancerous ovarian diseases.

EXAMPLES

Example 1: Gene Expression analysis

OSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the OSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), and differential expression in cancer tissue (CLASP 5). cDNA libraries were divided into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within

each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

(a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.

(b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.

(c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

CLASP Expression percentage levels for DEX0277 genes

DEX0279_25	SEQ ID NO: 25	PNS .0023	THR .0023	INL .0026	SYN .0028
DEX0279_26	SEQ ID NO: 26	PNS .0023	THR .0023	INL .0026	SYN .0028
DEX0279_27	SEQ ID NO: 27	KID .0006	LNG .0006	BRN .0008	TST .0011
DEX0279_28	SEQ ID NO: 28	KID .0006	LNG .0006	BRN .0008	TST .0011
DEX0279_30	SEQ ID NO: 30	FTS .0006	CON .0023	ADR .003	FAL .0063
DEX0279_31	SEQ ID NO: 31	FTS .0006	CON .0023	ADR .003	FAL .0063
DEX0279_35	SEQ ID NO: 35	INL .0038	SPL .0042	GLB .0046	CON .0102
DEX0279_36	SEQ ID NO: 36	INL .0038	SPL .0042	GLB .0046	CON .0102
DEX0279_39	SEQ ID NO: 39	BRN .0038	OVR .0082	LMN .0083	STO .0122
DEX0279_45	SEQ ID NO: 45	FAL .0063			

DEX0279_46	SEQ ID NO: 46		FAL .0063			
DEX0279_47	SEQ ID NO: 47	UTR .0075	SPL .0083	CRD .0091	BMR .0193	
DEX0279_48	SEQ ID NO: 48		THR .0091	BMR .0129	LMN .0139	
DEX0279_51	SEQ ID NO: 51		KID .0039	PLE .015		
DEX0279_53	SEQ ID NO: 53		CON .0011			
DEX0279_54	SEQ ID NO: 54		CON .0011			
DEX0279_55	SEQ ID NO: 55	GEM .0021	PNS .0022	LIV .0032	BLV .0037	
DEX0279_56	SEQ ID NO: 56	GEM .0021	PNS .0022	LIV .0032	BLV .0037	
DEX0279_57	SEQ ID NO: 57		NOS .0073			
DEX0279_58	SEQ ID NO: 58		NOS .0073			
DEX0279_65	SEQ ID NO: 65	GEM .0021	PNS .0022	LIV .0032	BLV .0037	
DEX0279_66	SEQ ID NO: 66	GEM .0021	PNS .0022	LIV .0032	BLV .0037	
DEX0279_67	SEQ ID NO: 67		MAM .0236	KID .027		
DEX0279_68	SEQ ID NO: 68		MAM .0236	KID .027		
DEX0279_72	SEQ ID NO: 72		UNC .012	UTR .0125		
DEX0279_77	SEQ ID NO: 77	TST .0027	BLD .0032	BLV .0033	PNS .0047	
DEX0279_78	SEQ ID NO: 78	INS .001	KID .0013	BLD .0032	INL .0032	
DEX0279_82	SEQ ID NO: 82		UTR .0075	PLE .0449		
DEX0279_83	SEQ ID NO: 83		UTR .0075	PLE .0449		
DEX0279_86	SEQ ID NO: 86		BRN .0004			
DEX0279_88	SEQ ID NO: 88		UNC .004	LIV .017		
DEX0279_90	SEQ ID NO: 90		OVR .001	ESO .0051		
DEX0279_91	SEQ ID NO: 91	INS .001	KID .0013	BLD .0032	INL .0032	
DEX0279_93	SEQ ID NO: 93		FAL .0063			

Abbreviation for tissues:
BLO Blood; BRN Brain; CON Connective Tissue; CRD Heart; FTS Fetus; INL Intestine, Large; INS Intestine, Small; KID Kidney; LIV Liver; LNG Lung; MAM Breast; MSL Muscles; NRV Nervous Tissue; OVR Ovary; PRO Prostate; STO Stomach; THR Thyroid Gland; TNS Tonsil / Adenoids; UTR Uterus

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard

curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the OSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the OSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the OSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 93 being a diagnostic marker for cancer.

Example 3: Protein Expression

The OSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the OSNA is subcloned in pET-21d for expression in *E. coli.* In addition to the OSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of OSNA, and six histidines, flanking the COOH-terminus of the coding sequence of OSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of OSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. OSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/1 of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, μg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al., Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

Based on the nucleotide sequences found by mRNA substractions the following extended nucleic acid sequences and amino acid sequences were determined.

DEX0279_1	DEX0126_1	DEX0279_94
DEX0279_2	DEX0126_2	DEX0279_95

DEX0279_3	DEX0126_3	DEX0279_96
DEX0279_4	DEX0126_4	DEX0279_97
DEX0279_5	DEX0126_5	DEX0279_99
DEX0279_6	DEX0126_6	DEX0279_100
DEX0279_7	DEX0126_7	DEX0279_102
DEX0279_8	DEX0126_8	DEX0279_104
DEX0279_9	DEX0126_9	DEX0279_105
DEX0279_10	DEX0126_10	DEX0279_106
DEX0279_11	DEX0126_11	DEX0279_107
DEX0279_12	flex DEX0126_11	DEX0279_108
DEX0279_13	DEX0126_12	DEX0279_109
DEX0279_14	DEX0126_13	
DEX0279_15	DEX0126_14	DEX0279_110
DEX0279_16	DEX0126_15	
DEX0279_17	DEX0126_16	DEX0279_111
DEX0279_18	DEX0126_17	DEX0279_112
DEX0279_19	DEX0126_18	DEX0279_113
DEX0279_20	DEX0126_19	DEX0279_114
DEX0279_21	flex DEX0126_19	DEX0279_115
DEX0279_22	DEX0126_20	DEX0279_116
DEX0279_23	DEX0126_21	DEX0279_118
DEX0279_24	DEX0136_1	DEX0279_120
DEX0279_25	DEX0136_2	DEX0279_121
DEX0279_26	flex DEX0136_2	
DEX0279_27	DEX0136_3	DEX0279_122
DEX0279_28	flex DEX0136_3	
DEX0279_29	DEX0136_4	DEX0279_123
DEX0279_30	DEX0136_5	DEX0279_124
DEX0279_31	flex DEX0136_5	DEX0279_125
DEX0279_32	DEX0136_6	DEX0279_126
DEX0279_33	DEX0136_7	DEX0279_127
DEX0279_34	DEX0136_8	DEX0279_128

DEX0279_35 DEX0136_9 DEX0279_130
DEX0279_36 flex DEX0136_9
DEX0279_37 DEX0136_10
DEX0279_38 flex DEX0136_10
DEX0279_39 DEX0136_11 DEX0279_131
DEX0279_40 flex DEX0136_11 DEX0279_132
DEX0279_41 DEX0136_12 DEX0279_133
DEX0279_42 flex DEX0136_12
DEX0279_43 DEX0136_13 DEX0279_134
DEX0279_44 flex DEX0136_13
DEX0279_45 DEX0136_14 DEX0279_135
DEX0279_46 flex DEX0136_14 DEX0279_136
DEX0279_47 DEX0136_15 DEX0279_137
DEX0279_48 flex DEX0136_15 DEX0279_138
DEX0279_49 DEX0136_16 DEX0279_139
DEX0279_50 flex DEX0136_16
DEX0279_51 DEX0136_17 DEX0279_140
DEX0279_52 flex DEX0136_17 DEX0279_141
DEX0279_53 DEX0136_18 DEX0279_142
DEX0279_54 flex DEX0136_18
DEX0279_55 DEX0136_19 DEX0279_143
DEX0279_56 flex DEX0136_19
DEX0279_57 DEX0136_20 DEX0279_144
DEX0279_58 flex DEX0136_20
DEX0279_59 DEX0136_21 DEX0279_145
DEX0279_60 DEX0136_22 DEX0279_146
DEX0279_61 flex DEX0136_22
DEX0279_62 DEX0136_23 DEX0279_147
DEX0279_63 DEX0136_24 DEX0279_148
DEX0279_64 flex DEX0136_24
DEX0279_65 DEX0136_25 DEX0279_149
DEX0279_66 flex DEX0136_25 DEX0279_150

DEX0279_67 DEX0136_26 DEX0279_151

DEX0279_68 flex DEX0136_26

DEX0279_69 DEX0136_27 DEX0279_152

DEX0279_70 DEX0136_28 DEX0279_153

DEX0279_71 flex DEX0136_28

DEX0279_72 DEX0136_29 DEX0279_154

DEX0279_73 DEX0136_30 DEX0279_155

DEX0279_74 flex DEX0136_30

DEX0279_75 DEX0136_31 DEX0279_156

DEX0279_76 DEX0136_32 DEX0279_157

DEX0279_77 flex DEX0136_32 DEX0279_158

DEX0279_78 DEX0136_33 DEX0279_159

DEX0279_79 flex DEX0136_33

DEX0279_80 DEX0136_34 DEX0279_160

DEX0279_81 flex DEX0136_34

DEX0279_82 DEX0136_35 DEX0279_161

DEX0279_83 flex DEX0136_35

DEX0279_84 DEX0136_36 DEX0279_162

DEX0279_85 DEX0136_37 DEX0279_163

DEX0279_86 DEX0136_38 DEX0279_164

DEX0279_87 flex DEX0136_38

DEX0279_88 DEX0136_39

DEX0279_89 flex DEX0136_39

DEX0279_90 DEX0136_40 DEX0279_165

DEX0279_91 DEX0136_41

DEX0279_92 flex DEX0136_41

DEX0279_93 DEX0136_42 DEX0279_167

The follow chromosomal locations were determined.

DEX0279_1 chromosome 1

DEX0279_3 chromosome 3

DEX0279_4 chromosome 11

DEX0279_7 chromosome 14

DEX0279_11 chromosome X
DEX0279_12 chromosome 9
DEX0279_13 chromosome 3
DEX0279_21 chromosome 12
DEX0279_22 chromosome 16
DEX0279_23 chromosome 2
DEX0279_26 chromosome 17
DEX0279_27 chromosome 12
DEX0279_29 chromosome 8
DEX0279_31 chromosome 10
DEX0279_40 chromosome 3
DEX0279_45 chromosome 10
DEX0279_46 chromosome 10
DEX0279_48 chromosome 14
DEX0279_50 chromosome 2
DEX0279_52 chromosome 11
DEX0279_57 chromosome 16
DEX0279_58 chromosome 16
DEX0279_59 chromosome 19
DEX0279_62 chromosome 9
DEX0279_63 chromosome 10
DEX0279_69 chromosome 10
DEX0279_71 chromosome 2
DEX0279_77 chromosome X
DEX0279_78 chromosome 8
DEX0279_79 chromosome 8
DEX0279_83 chromosome 2
DEX0279_84 chromosome 11
DEX0279_88 chromosome 12
DEX0279_90 chromosome 1
DEX0279_91 chromosome 8
DEX0279_92 chromosome 8

The following Jamison-Wolf antigenic sites were also determined.

DEX0279_94 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
47-84	1.10	38
32-44	1.04	13

DEX0279_96 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
45-58	1.09	14

DEX0279_98 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
38-51	1.17	14

DEX0279_99 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
56-71	1.12	16
15-44	1.10	30

DEX0279_100 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
61-72	1.20	12
15-44	1.14	30

DEX0279_101 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
85-98	1.12	14

DEX0279_102 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
14-27	1.24	14

DEX0279_107 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
15-24	1.19	10

DEX0279_108 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
575-585	1.30	11

	322-336	1.21	15
	415-425	1.12	11
	889-916	1.08	28
	373-389	1.07	17
	832-876	1.04	45
	757-815	1.03	59
	1018-1035	1.02	18
	677-698	1.01	22
DEX0279_110		Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	69-83	1.21	15
DEX0279_112		Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	60-74	1.16	15
DEX0279_115		Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	259-271	1.15	13
	204-216	1.07	13
	391-401	1.07	11
	587-653	1.04	67
DEX0279_116		Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	20-31	1.05	12
DEX0279_118		Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	30-42	1.12	13
DEX0279_119		Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	84-97	1.12	14
DEX0279_120		Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	88-117	1.26	30

119-142	1.24	24
31-66	1.07	36
70-84	1.03	15

DEX0279_121 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
55-75	1.14	21

DEX0279_124 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
62-72	1.15	11

DEX0279_125 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
9-55	1.03	47

DEX0279_126 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
32-46	1.12	15

DEX0279_127 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
37-51	1.13	15
16-33	1.08	18

DEX0279_138 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
69-80	1.07	12

DEX0279_141 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
2-11	1.01	10

DEX0279_142 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
2-32	1.20	31

DEX0279_144 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
34-62	1.03	29

DEX0279_147 Antigenicity Index(Jameson-Wolf)

	positions	AI avg length
	17-27	1.30 11
DEX0279_148		Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	15-51	1.01 37
DEX0279_150		Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	50-59	1.10 10
	166-183	1.03 18
	109-160	1.00 52
DEX0279_154		Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	13-34	1.16 22
DEX0279_157		Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	24-37	1.09 14
DEX0279_160		Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	24-38	1.00 15
DEX0279_164		Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	22-31	1.08 10
DEX0279_166		Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	74-83	1.03 10
DEX0279_167		Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	12-44	1.26 33

In addition, the following helical regions were predicted.

DEX0279_106 PredHel=5 Topology=o10-32i44-66o81-103i110-132o136-158i

DEX0279_109 PredHel=1 Topology=i45-67o

DEX0279_110	PredHel=1	Topology=i43-65o
DEX0279_125	PredHel=1	Topology=i93-115o
DEX0279_132	PredHel=6	Topology=o4-21i68-85o100-122i153-172o182-201i222-239o
DEX0279_135	PredHel=2	Topology=i21-43o53-75i
DEX0279_159	PredHel=1	Topology=i7-29o
DEX0279_161	PredHel=1	Topology=i13-35o
DEX0279_163	PredHel=1	Topology=o15-34i

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 93. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al., Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al., Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 μg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater.

Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid,

aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a

pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 μg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf

intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 1. Preferably, the 5'primer contains an EcoRI site and the 3'primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the

introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach,

intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell.

Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon,"Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be

accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions

of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or

vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 94 through 167;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 93;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 60% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.

6. A method for determining the presence of an ovary specific nucleic acid (OSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule according to claim 1 under conditions in which the nucleic acid molecule will selectively hybridize to an ovary specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to an OSNA in the sample, wherein the detection of the hybridization indicates the presence of an OSNA in the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 94 through 167; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 93.

12. An antibody or fragment thereof that specifically binds to the polypeptide according to claim 11.

13. A method for determining the presence of an ovary specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the ovary specific protein; and

(b) detecting binding of the antibody to an ovary specific protein in the sample, wherein the detection of binding indicates the presence of an ovary specific protein in the sample.

14. A method for diagnosing and monitoring the presence and metastases of ovarian cancer in a patient, comprising the steps of:

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the ovary specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of ovarian cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with ovarian cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the ovarian cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

SEQUENCE LISTING

<110> Salceda, Susana
Macina, Roberto
Recipon, Herve
Cafferkey, Robert
Sun, Yongming
Liu, Chenghua
diaDexus, Inc.

<120> Compositions and Methods Relating to Ovarian Specific Genes and Proteins

<130> DEX-0280

<150> 60/252,061
<151> 2000-11-20
<150> 60/253,257
<151> 2000-11-27

<160> 167

<170> PatentIn version 3.1

<210> 1
<211> 670
<212> DNA
<213> Homo sapien

<400> 1
```
gcgtggtcgc ggcgaggtcg tgcctgtaat cccagctact tggaagtctg aggcaggaga      60

atcgcttgaa tctgggagtt ggaggttgca gtaagccaag atcgcgccac agcactccag     120

cctaggtgac agagtgacac tctgtctcaa aaaaaaaaaa aaaaaaagag aggggtctc     180

cgggccgggc agggtgggct cacgcctggt gaatcccagc aattttgggg agggccgagg     240

tggggcggga tcacttggag atcgggagtt tggagaccag tctgtgccaa cttgtgtgga     300

aaccccgtct ctacttaaaa ttggcaaaat tagctaggct tgggtgggca catggcctgt     360

aatctcagct actggggacc ctggagggcg gcgaaatcta tttggaccct ggtggaggtg     420

ggcgggttgg cagtcagctt ggactgttgg ccaccacggc actccaagcc tggggcagca     480

gaggggagac tcctttccac aaaaaaaaaa aaaaaaaaaa aaaacggggg ggggtgtacc     540

cgggggcgca aaagggggtg ccgggggggg aatggggtgt tccgggcgcc aaattccccc     600

cacattctcg cgaaggaaaa gtgtaaaaga aaaaaaaaga gaaaaagaaa aagaaaagag     660

aaaagaaagc                                                            670
```

<210> 2
<211> 1082
<212> DNA

<213> Homo sapien

<400> 2

```
gggtcgcggc cgaggtacag actcaagcgg gccaacaaga tagatcaaaa ccactgatgt     60
gagaagataa gtccttgaag caggcaattg cccttaaaag cagtcatccc tgcccctgtg    120
gatcggcctg gatgccatct gccctacacg tagatctgtt ggctgtgtgg ccactgaccc    180
gccccccatc ctcctgcctc gctctctcca cagtgtttgt acgctgatga ctggtttctg    240
tagtacgatc ctgtggtcta gcggcgtcaa taggtgctac catctccctt ggaatcctag    300
caagagcccc cgtacacgtc tcagaggacg ccgtacttgc ctgtaggagt tctaggcgcg    360
tggtggccgc tgtactggtg tccgggaggc cgaggcctgc agcgatgttc ccacggggtc    420
acccccgcag gaggacgcac aggcgatcgt ggaagtgctc tctgcgccac ttctgggtgc    480
ttgtactggt cttcttggcg ttgctttccc aatcaccttg gcaacatcac ttaggatgaa    540
ttggtacgct gtttattagg tcctctgata atagccgtat aagacacttt gtgactgtta    600
caatatatgg gaaggtttga ctgttgggtt atgctgtact tgaagaggat ctgaagacgt    660
ggtcgaatgt ggttcaatca tttcagccgt cgctaataaa ttgtagagga caatgcagtg    720
gtcaagggta gtggtaggca tgcgcctgta atctcagctg ggcgagaca cgcgggagcg     780
acaggagatc cgcgggggga acccggggag gcgggggaaa gtgcgagcaa aaaaaaccaa    840
aaacgggggc caaatctggg aaaacggatg gaaacgggtt caaaaccgaa cacacacaca    900
aacccctaaa aaggttaggg atataagggg aaccagggga cacacgtttc acccaaagct    960
aaacaattga acctaagtgc gtctttgcat gtttaatacc agatctaaca ggttataatg   1020
gacattccca gaaaacaaac aaaaggctgt gtgcaggcat agattcggaa gagtacgcac   1080
aa                                                                  1082
```

<210> 3
<211> 691
<212> DNA
<213> Homo sapien

<400> 3

```
cgagcggccg cccggggcag gtacggctat tggttggtgg caaatgtaaa ttagcgcaac     60
cactatggag aacagttcgg catggatctt caacaaaact aaaaatatga gctaccattg    120
tgatccagca attccactgc taggtatata cccaagaagt aagaggaaat tagatgtgga    180
agagatgtct gcgctcttat gttgattgca gcactgttca caatagccaa gattgggaag    240
caatgtaagt gtctaccaat cttgacgaac ggtataaatg gaagaggtgg ggcctgggcg    300
tggtggctca tgcctgtaga tcccagcact ttgggaggcc gaggtcagat cacctgaggt    360
```

cagaagtttg agaacagcct ggccaatatg gagaaacccc atctttacta aaatacaaaa 420

attagctggg cgtggtggcg cacacctgta ttcccagcta ctcgggaggc tgaggcagga 480

caattgcttg aacctgggag gcagagatag cagggagcca agattgtggg catacagagc 540

aagactccct ctcaaaaagg agtccagtcc tgggtgaaca aaagtgaaga ctcccatctc 600

gcggggggt aaccggggcc aaaagggtcc cggggggaaa atggggtccc gccccaaacc 660

ccccgaaaga aaccaagcaa ccaaaaaaga a 691

<210> 4
<211> 953
<212> DNA
<213> Homo sapien

<400> 4
aggtgatcat cactataggg ccactggttc atctagatgc tgctcgagcg gcgcagtgtg 60

atggatggcc gcccgggcag gtgtgaaacc ccatctctaa taaaatacaa aaattagctg 120

ggtgtggtgg cgggcgcctg tagtcccagc tactggggag gctgagacag gagagtggca 180

tgaacccggg aggcggagct tgcagtgagc cgagttcgca ccactgcact ccagcctggg 240

caacagagcg agactctgtc tcaaaaaaaa acaaaagaaa aactcaataa cttgggccaa 300

gctacttaac atctccaggc atcaattcca ttacctggaa gataagcaaa acagtacata 360

cccaatacag ttgcaaaact aaattaaata atttaagcaa agctctcaat attggacctg 420

gcataaagtt agtaccaata aatgttagct actattatct ttcataataa gtgaaatcat 480

aggctacata ttatttgaaa gaagaaaaat atacatgtgt ataaattctt ccatttaaaa 540

tagttaaaag tgtttgctct tgaaaacttt aaagactacc ttttggaaga ttcactttta 600

taatatttta ctaagagaag taaataaggt taattatatt cttcaaagac aaatgtgtat 660

gcccctcaag aaatcacaca acggcattcc aaggttaaat ttttaccatg tttgggtatc 720

aaattctatc tggcttgggc atgttattat tgtgaaccaa tctcttactc atccataggc 780

cataccttgg aatattgtgg tctgggtccc agactgctgt tatacaagtg aatggtatca 840

gtaaagtgga ttcctaaata ggggcaagtg ctgcaaattt ctgggttccc agaggaagat 900

aaagtttgtg ggcaaattca aaaccccaaa gggggacaag cccttgttca aca 953

<210> 5
<211> 434
<212> DNA
<213> Homo sapien

<400> 5

gcggccgccc gggcaggtac gtgtcgcaag catcctcgca cgatcccgag agcccgtgga 60

gcgggggctt gctgggccgt cgcactcatt tacccgggcg accgtgcgag gctcgttctg 120

cgttgtcgtg gtggtgcaga gcctcgagtg catcatggag tgctgagatc gaggtgcccc 180

cttgtggcac ctgctctggt ccacggtgag gctggctata gagggaagca aggagccgtc 240

gcgcagtcca gctcgcggag gcggtgcgtc tagctaggta ggtatcctcc ggactgcccg 300

gttgctcgtc ccatgtcctg gggtatcgtc ctggcgatga agccttgaca ggcaggtagg 360

ctggcctggt tctgtgtccg gtcattcccg ggatgggggc agcggtgtac ctcggccgct 420

accaacgcta agcc 434

<210> 6
<211> 703
<212> DNA
<213> Homo sapien

<400> 6
cgagcggcgc cgggcaggta cccaacagct cattgagaac gggccaggat gacgatggcg 60

gctttgtgga atagaaaggc gggaaaggtg gggaaaagat tgagaaatcg gatggttgcc 120

gtgtctgtgt ggaaagaagt agacatggga gacttttcat tttgttctac actaagaaaa 180

attcctctgc cttgggatcc tgttgatctg tgaccttacc cccaaccctg tgctctctga 240

aacatgtgcg gtgtccactc agggttaaat ggattaaggg cagtgcaaga tgtgctttgt 300

taaacagatg cttgaaggca gcatgctcgt taagagtcat caccaatccc taatctcaag 360

taatcaggga cacaaacact gcggaaggcc gcagggtcct ctgcctagga aaccagaga 420

cctttgttca cttgtttatc tgctgacctt ccctccacta ttgtcccatg accctgccaa 480

ataccccctct gtgagaaaca cccaaggaat tatctaaaaa aaaaaaacaa aaacacacaa 540

caacaaaaaa agcgcttggg ggggaaccaa ggggccaaag gggggtcccc cggggggac 600

cagtggggtt ccccggccca caaattcccg cccaaaatag ggggtacaac aggaagacac 660

aaacaacagg caaaaacaca aaaagaacca gcacaaacaa aaa 703

<210> 7
<211> 823
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (195)..(195)
<223> a, c, g or t

<400> 7
gcggggcccg ggcaggttgt aatcccagct acttgggagg ctgaggcaga gaattgcttg 60

aacccgggag gcagaggttg cagtgagtcg agatcgtacc actgcactcc agccaggcaa 120

cagaaggaga ctccatctca aaaaaaagaa aaaaaggtaa ggccggactc agtggctcac 180

acttgtaatc tcagncactt cgggaggagg ctgaggcagg cagattgctt gcgcttagga 240

gttcaggact gaactaggca acatggagaa accatgtctc tacaaaatat aaaaaaatta 300

gctggacatg gtgtcttgca cctgtagtcc cagctactca ggaggctgag ctgggagtat 360

cacttgagcc caggaagtgc agattgcagt agccaagatc atgccactgc actccagcct 420

gggaaacata gtgagatcct gtctcaaaaa taataataat aaaataggcc gagcgcggtg 480

gctcacgcct gtaatcccag cactttggga ggccaaggcg ggtggatcac gaggtcagga 540

gatcaagacc atcctggcta acacggtgaa accccatctc tactaaaaat acaaaaaatt 600

tagcccggtg tggtggtggg cgcctgtagt ccccagctac tagggaggcg gaggcaggag 660

aatggcgtga acccgggagg tggagcttgc agtgagccga gattgcacca ctgcactcca 720

gcctgggtaa tacagcgaga ctccattcca aaaaaaaaaa acaaagaaaa aaagctgggg 780

gtacctggca aagtggtccg ggggaattgt tcgtcacccc agg 823

<210> 8
<211> 327
<212> DNA
<213> Homo sapien

<400> 8
cacaacatac gagcaatacg agcagggtgt agacgacggc acgggaatga cgccgtcatg 60

ggtgcagaaa tccggcatgt ctagtcatac tgcctgagaa gatgccccga tggagcgaag 120

tatgagactg cgggtggcag ggccattgag gacacgcaat ggtcagaatc aactattgaa 180

gagctcgtcc gcatggtgct agaactaggg tgaggaacct ttcccaagtc tccccagagt 240

ccggtgcctt tacccgtgtg gtgaaattat gtatctagcc taaaggtaat cccctagagc 300

tttatcttgt ctcacgcaaa ttttcac 327

<210> 9
<211> 683
<212> DNA
<213> Homo sapien

<400> 9
attccacacg aacactacgg aatcaaatac gatacaccca gagccgccac aggacgcact 60

caccaaaccc atggagaaac cacaccacgc cctatcccac aagaaacaaa acacacacca 120

cgacgacaca caccccacgg ccccccacac caacccacac caagcaacga cccagcacaa 180

cacaaacaac cacacacacc acaagatgac gcgcaaaaca cacacagagc agaccaacac 240

ggcccaccca caacgagtca gtgcaaaagt gtgagctcag gcgagacgac acaacgaatg 300

caatcgacag ccacaatgat atacaaacac acgatcaaac tccacgacgg ccacgctcac 360

aacgaccaca aggcgaagaa gaaagacagc cacaaccaga cgagggcgaa gaaagaaggg 420

aaggctgtga gcgcggaggg ttgcctgaga cagagcggtc gtgactgacc agcacggggc 480

ggggatggag tgtagcggaa agggcgcgag gtgaagcgga gcgggcagaa gccctgagcg 540

cgagagccgc gaacgacgtt acaccacgag ccgagaggaa agcagtgaca agcactccag 600

gcacatagac agcaccaaga tgcccacaaa agaggccact aaaagcaagc aaaacgaaga 660

aacgctagaa ggtcgcagaa ccc 683

<210> 10
<211> 1286
<212> DNA
<213> Homo sapien

<400> 10
taaacacaaa gatgactaga tcgactcact ataggccgct tgtttcactc tagaatacat 60

gctcgaccgg ccgccaatgt gatggatctg gccgcccggg caggtaaaac caggcctggc 120

taattttgaa attattgtaa gcgatgggtc tagctgtgtt tgcacatgct aggtctcaaa 180

ctacctggcc tacaagacat accctgccct gcctctacgc ctatggcctc ctaaagtgat 240

gtgtgattac agtgactgtg ggccaccaat gcccagcctg aatctataat ttattattat 300

ttgggagatt atagatattg ttaaattttt aaaacaagac aatcggaaag gttataagaa 360

gtgacacgcc atcactcact aagccactca caaacgtgca ccccctcaaa caggacctgg 420

caatgccaaa aagggaacgt gccacgtgca cctcaaataa ggttaaacca gtggtgtgag 480

gggcaaaata tgagaagcaa ggggtgacac tagagaggaa caaaatggat gtatccacaa 540

tgagacccga gattataaca ctaaaagcgg gaaataagaa caacgcagaa caaacaaaac 600

acaagacgaa acacacaaaa cagtacagag aaaaaaaaca agaagagaaa cacgaggcga 660

ccgcccgagg agaagaaaaa ggaaaaacag ggacacaaaa agacaccaag aagataaagc 720

atttatatgc ataatggggg agaaagaaaa tggaaaagcg ggaagttaaa atactgaatg 780

aataatgatt tagttagtga aaggatgcat aagagagagt gcacataggg agagagatag 840

gtcgataaag aaaaaaacaa acgcagaaaa ggaccaaaag agagcggaga acagaaggac 900

aaaaacagaa agaagggcca aaaaaagcgg gaaaaaaaaa aaggaaaggc acgcacaaca 960

gatccgaccg tcagctacag gggagcaaag aaagaagaaa acgggaggaa aaagagcgac 1020

agacgagcac agagagccca acgaacgacc caagcggagg acagcaacac aggagaccag 1080

acaagagaac gagaaaaagc aacaaagaga gagacaaaca aaagacaaga agaaaaaaga 1140

agaaaaagcg aaacaaaaag aaaacacacg gacgacggcg cgccagaaag aaaaaacacg 1200

aaaaaagaac agaaaggagc aagagacaac aaaagaacaa aaaaaaacca aaagaaaacg 1260

aaaaaagagg acaaaacccg gcaggg 1286

<210> 11
<211> 739
<212> DNA
<213> Homo sapien

<400> 11
tcgcggccga ggtctcgtga gccccctaga ccatcacgga tgccgagctt cgggtaactc 60

tcacagtgga aggttcccac gccgccccta atcccgctcg aagcagccct gagaaacatc 120

gcccattctc tctccatatc acccccaaa aatttttgcc accccaacac ttcaacacta 180

tttgttttat ttttcttatt aatataagac ggcaggaatg tcaggcctct gagcccaagc 240

caagccatcg catcccctgt gacttgcacg tatatgccca gatggcctga agtaactgaa 300

gaatcacaaa agaagtgaat atgctctgcc ccaccttaac tgatgacctt ccaccacaaa 360

agaagtgtaa atggccggtc cttgctttaa gtgatgacat taccttgtga aagtcctttt 420

cctggctcat cctggctcaa aaatcatccc cactgagcac cttgcaaacc cccactcctg 480

cctgccagag aacaaacccct ctttgactgt aattttcctt tacctaccca aatcctataa 540

aacggcccca cccttatctc ccttcgccga ctctcttttc ggactcagcc cgcctgcacc 600

caggtgaaat aaacagccac gttgctcaca aaaaaaaaaa aaaaaaaaaa aagggccggg 660

ggaaccgggg ccaaaagcgt cccggtggga atggttcccc gccccaatca cccaaaaaaa 720

aaaggaggaa aaaaaaagc 739

<210> 12
<211> 4640
<212> DNA
<213> Homo sapien

<400> 12
atggccttgg ctgtcccgtg tgtgacctgg atgtcatggt gccacctcct tcctgggagc 60

agtaaggatt ccatgccatg gaggacagag agcttgggct gcagggatgg atgcagcctc 120

tgctttacat cccaccccgt ggagaagctc ccagggaagc ggggagggtc atggctgccc 180

aggtcccatc tgatgccgag gctgaaggag tctcgctccc acgagtccct gctcagcccc 240

agcagtgcgg tggaggcgct ggacctcagc atggaggaag aggtggtcat caagcccgtg 300
cacagcagca tccttggcca ggactactgc ttcgaggtga cgacgtcatc aggaagcaag 360
tgcttttcct gccggtctgc agctgagcgg gataagtgga tggagaacct ccggcgagcg 420
gtgcatccca acaaggacaa cagccggcgt gtggagcaca tcctgaagct gtgggtgatc 480
gaggccaagg acctgccagc caagaagaag tacctgtgcg agctgtgcct ggacgatgtg 540
ctctatgccc gcaccacggg caagctcaag acggacaatg ttttctgggg cgagcacttc 600
gagttccaca acttgccgcc tctgcgcacg gtcactgtcc acctgtaccg ggagaccgac 660
aagaagaaga agaaggagcg caacagttac ctgggcctgg tgagcctacc tgctgcctcg 720
gtggccgggc ggcagttcgt ggagaagtgg tacccggtgg tgacgcccaa ccccaagggc 780
ggcaagggcc ctggacccat gatccgcatc aaggcgcgct accaaaccat caccatcctg 840
cccatggaga tgtacaaaga gttcgctgag cacatcacca accactacct ggggctgtgt 900
gcagccctcg agcccatcct cagtgccaag accaaggagg agatggcatc tgccctggtg 960
cacatcctgc agagcacggg caaggtgaag gacttcctga cagacctgat gatgtcagag 1020
gtggaccgct gcggggacaa cgagcacctc atcttccggg agaacacact ggccaccaag 1080
gccattgagg agtacctcaa gctagtgggc cagaagtacc tgcaggacgc cctaggtgag 1140
ttcatcaaag cgctgtatga gtcagatgag aactgcgaag tggatcccag caagtgctcg 1200
gccgctgacc tcccagagca ccagggcaac ctcaagatgt gctgcgagct ggccttctgc 1260
aagatcatca actcctactg tgtcttccca cgggagttga aagaggtgtt tgcctcgtgg 1320
aggcaggagt gcagcagtcg cggccgcccg gacatcagtg agcggctcat cagcgcctcc 1380
ctcttcctgc gcttcctctg cccagccatc atgtcgccct cactcttcaa cctgctgcag 1440
gagtaccctg atgaccgcac tgcccgcacc ctcaccctca tcgccaaggt cacccagaac 1500
ctggccaact ttgccaaatt tggcagcaag gaggaataca tgtccttcat gaaccagttc 1560
ctagagcatg agtggaccaa catgcagcgc ttcctgctgg agatctccaa ccccgagacc 1620
ctctccaata cagccggctt cgagggctac atcgacctgg gccgcgagct ctccagcctg 1680
cactcactgc tctgggaggc cgtcagccag ctggagcaga gcatagtatc caaactggga 1740
cccctgcctc ggatcctgag ggacgtccac acagcactga gcaccccagg tagcgggcag 1800
ctcccaggga ccaatgacct ggcctccaca ccgggctctg gcagcagcag catctcagct 1860
gggctgcaga agatggtgat tgagaacgat ctttccggtc tgatagattt cacccggtta 1920
ccgtctccaa cccccgaaaa caaggacttg ttttttgtca caaggtcctc cggggtccag 1980

ccctcacctg cccgcagctc gagttactcg gaagccaacg agcctgatct tcagatggcc 2040
aacggtggca agagcctctc catggtggac ctccaggacg cccgcacgct ggatggggag 2100
gcaggctccc cggcgggccc cgacgtcctc cccacagatg ggcaggccgc tgcagctcag 2160
ctggtggccg ggtggccggc ccgggcaacc ccagtgaacc tggcagggct ggccacggtg 2220
cggcgggcag gccagacacc aaccacacca ggcacctccg agggcgcgcc aggccggccc 2280
cagctgttgg caccgctctc cttccagaac cctgtgtacc agatggcggc tggcctgccg 2340
ctgtcacccc gtggccttgg cgactcaggc tctgagggcc acagctccct gagctcacac 2400
agcaacagcg aggagttggc ggctgctgcc aagctgggaa gtttcagcac tgccgcggag 2460
gagctggctc ggcggcccgg tgagctggca cggcgacaga tgtcactgac tgaaaaaggc 2520
gggcagccca cggtgccacg gcagaacagt gctggccccc agaggaggat cgaccagcct 2580
ccgcccccac ccccgccgcc acctcctgcc ccccgcggcc ggacgccccc caacctgctg 2640
agcaccctgc agtacccaag accctcaagc ggaaccctgg cgtcggcctc acctgattgg 2700
gtgggcccca gtacccgcct gaggcagcag tcctcttcct ccaaggggga cagcccagaa 2760
ctgaagccac gggcagtgca caagcagggc ccttcacctg tgagccccaa tgccctggac 2820
cgcacagccg cttggctctt gaccatgaac gcgcagttgt tagaagacga gggcctgggc 2880
ccagaccccc cccacaggga taggctaagg agtaaggacg agctcagcca agcagaaaag 2940
gacctggcgg tgctgcagga caagctgcga atctccacca agaagctgga ggagtatgag 3000
accctgttca agtgccagga ggagacgacg cagaagctgg tgctggagta ccaggcacgg 3060
ctggaggagg gcgaggagcg gctgcggcgg cagcaggagg acaaggacat ccagatgaag 3120
ggcatcatca gcaggttgat gtccgtggag gaagaactga agaaggacca cgcagagatg 3180
caagcggctg tggactccaa acagaagatc attgatgccc aggtatacac agccctaagg 3240
agcctgtccc atgacccccg ctcacatccc cattgtccac aggagaagcg cattgcctcg 3300
ttggatgccg ccaatgcccg cctcatgagt gccctgaccc agctgaaaga gaggaaagtg 3360
tcatccgtgc tgctcacagc gggcacagct gccaagcctt ccctgctcat cctccacatc 3420
agcactggtc ctcagaccga ccagcccaag aaacatctca ccaatttcaa atccgcctgt 3480
gtcctcaaga acttaaaacc tcttcaactc acacctgacc taaaacctaa acgccttatt 3540
ttcttctgca acaccgcttg gccccaatac aaactcgaca atggctctaa atggccagaa 3600
aacggcactt tcgatttctc catcctacaa gacctaaata atttttgtcg aaaaatgggc 3660
aaatggtctg agatgcacag gaaagagttt ttttctctag cccaatctca tgctgataac 3720
cgccggcttc atgagccaga cctccaggaa ggcattagag cagttccccg agaggatccc 3780

caatggaact accaggcaaa ttccccaggt atagctaagc aagattacat ggtttcctgc 3840

gtagttgaag ggcttaaaaa agcagcttac aaagctatta attatgacaa acttaaagaa 3900

ctacccaaga tcgcctcgtg agccccctag accatcacgg atgccgagct tcgggtaact 3960

ctcacagtgg aaggttccca cgccgcccct aatcccgctc gaagcagccc tgagaaacat 4020

cgcccattct ctctccatat cacccccaa aaatttttgc caccccaaca cttcaacact 4080

attttgtttt atttttctta ttaatataag acggcaggaa tgtcaggcct ctgagcccaa 4140

gccaagccat cgcatcccct gtgacttgca cgtatatgcc cagatggcct gaagtaactg 4200

aagaatcaca aaagaagtga atatgctctg ccccacctta actgatgacc ttccaccaca 4260

aaagaagtgt aaatggccgg tccttgcttt aagtgatgac attaccttgt gaaagtcctt 4320

ttcctggctc atcctggctc aaaaatcacc cccactgagc accttgcaac ccccactcct 4380

gcctgccaga gaacaaaccc tctttgactg taattttcct ttacctaccc aaatcctata 4440

aaacggcccc acccttatct cccttcgctg actctctttt cggactcagc ccgcctgcac 4500

ccaggtgaaa taaacagcca cgttgctcac aaaaaaaaaa aaaaaaaaaa aaagggccgg 4560

gggaaccggg gccaaaagcg tcccggtggg aatggttccc cgccccaatc acccaaaaaa 4620

aaaaggagga aaaaaaaagc 4640

<210> 13
<211> 760
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (602)..(602)
<223> a, c, g or t

<400> 13
cgccgatcga agatagaccc tatgggcgat gtgcctctag atgctgctcg agcggcgcca 60

gtgtgatgga tcggccgccc gggcaggtac ggtattggtt ggtggaaatg taaattagca 120

caaccactat ggagaacagt ttggaggatc ttcaaaaaac taaaaataga gctaccatat 180

gatccagcaa ttccactgct aggtatatac ccaaaagaaa ggaaattaga tgtggaagag 240

atgtctgcac tcttatgttt attgcagcac tgttcacaat agccaagatt tggaagcaat 300

gtaagtgtct accaacagac gaacggataa agaaaaggtg gggccgggcg tggtggctca 360

tgcctgtaat cccagcactt tgggaggccg aggcagatca cctgaggtca gaagtttgag 420

aacagcctgg ccaatatgga gaaaccccat ctttactaaa atacaaaaat tagctgggcg 480

tggtggcgca cacctgtagt cccagctact cgggaggctg aggcaggaga attgcttgaa 540

cctgggaggc agagattgca gtgagccaag attgtgggca cagagcaaga ctccctctca 600

anaaaggagt aaattatacc aaaaccagca aaaaaaaaaa aaaagcctct gggggaaccc 660

ggggcccaaa gctgttcccg gtgtgaaatt ttttcccgcc cacattccca caatgcacaa 720

acacaaaatt cgcaaatgaa acctaaaagg atggaacaaa 760

<210> 14
<211> 604
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (414)..(414)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (416)..(416)
<223> a, c, g or t

<400> 14
tttttttttt tttttttttt ttttcttttt tgggagtctt gccctgtaac cccagctact 60

ggggaaactg aggtaggaaa gtcgcttgag gtaggaaaat cgcttgaatg gctttgagct 120

gaaatcatgc cactgcactc tacccggggt aacaaagcaa gactccatca aaaaaaagag 180

aagagagaaa agaaaagttg ggtaattatt tagggaaagc tatacagaat aagtagaagt 240

tggccaggtg aagtcgggga ggatgattat ttaagcagaa agaaatattt acccagtatg 300

gtgtgttggg cagagggaca cttctcacga ggaacacctg tatggaggct gccccctgtg 360

gcgggaaaga aacccaccgt gaacagcggc tgtgaaaaag cgtgggggga tacngnccat 420

gtaggctcat taaggctgtg ctttctcccg tgggtggtgg aaagatgtgt gcgttatctc 480

cggctctcca caaattcctc caccaacaac cattccgcgc caaggaacca aaagggtgga 540

aagaagcaac caaaccaaga agacagaaaa agagaaaaca aaaacaaaaa aacaaaaaaa 600

aggg 604

<210> 15
<211> 974
<212> DNA
<213> Homo sapien

<400> 15

ggcctggccg ggcttgtggg agccgtcagc agcgcgaggt gctgcgaggc gcaatgctta 60

gaattactat attcagcgcg tttgcttttg tgcgcatgaa tcccttggag ctggggggtc 120

aagtcaggct gaaagaacac tttgcagtga tcggccttgt cacaaggtgc ataaaccttt 180

tacacggaga gtgtctttca cagaggtatg ggatcgaggt ccacgttgcg tctctcagga 240

acacgaatcg gtaccttccg tgcacatgca cctaaacaga gcagtgcaat cctcgttact 300

gactaggatt cgtgcagata taccggttcc tagggtttcg ctacatatca tagggtccat 360

catacttccc ggcaaatagc atttcttgat tgcactgaaa ctactgcata gagtagttct 420

tgggaggaat ccttgcatgc gcgtactgta agtatagggg ttcgatccat ccaactttga 480

ttaaatgatg ccaattgcat tcctgctcgt atgaaggtca gaagtagctc gggtggcgcc 540

cttgctccag gtgcgtatac agatggcgta ccatcgctgg cggtgtagaa aaagcctcgt 600

ttgctccgct tcgtctcctc tcttgttctc tccatgaaag tcattttggg ggcgggaaaa 660

attgggttgg cgttcggtaa ttccaatggg tttacttcct tggttggttg aaaacaggtc 720

tcggtatata tcccgttttt cctactatga gattttcccc ggctagcaaa gcaatacggg 780

aggagccaag gagaacacac agaaacagag ggaaaaaaaa aatagaaaga aaagaaaaca 840

aaaacaagcg cgcgtggtgt ggggggtac acacaccgtg gtggtgccaa aatgggggtt 900

tttccccgt gtggggtggt gagaaaatgt gttatttcac cccgggctcc ccacaattct 960

cccccccgga acaa 974

<210> 16
<211> 863
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (93)..(93)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (222)..(222)
<223> a, c, g or t

<400> 16
gcttttttttt tttttttttt ttttttggta aaggaaaaat aggcccgtta ttttttcctc 60

tggaccaact ggcacttctt tgaaacccct gcntgttgtt gccaaggccc tttcccccaa 120

agggagggat atattcaggt ggtgtggacg ccaatgctct ctcagggtgt gaaaaggagc 180

cctgtggcac caacaacaaa acgcagagac tctccaagag cncctcctct ccatatatgt 240

ggacgagaga tatctcacac gccgtggggg agataactct catgtggtct catatagcgc 300

gtgtgttccg cgtgtgtgtg tagaaatgtg tgtgtatatc tgccggtcac tcacacatat 360

ctctcacaca cacaacaaca catataagcg ggaaaaaaca aagttgccca agatatgagg 420

cacattcaat tccacacaaa acatatcacg cagtctaggt ggcacaagaa acatagatca 480

acgacatagt gcgacctata tcgacaactc gacaccaata ttatagcaac acacaatagg 540

gacaaagaga ggataccaat tgtgccattt gaaccccatt agcgaatatc cattatgagt 600

gatcaaaggc gatacacaat tccaacccag gcagatgaaa aaatagagaa gaggatgatc 660

aaaaagagag gagactgttt aaaccccсaa tacaagagac aaccatgtac aataaaccac 720

cgaatcaggg gacatgataa aaaatgatcc cacaggtaaa caatttatac ccacaagtcg 780

gacaacaacc aaaccttagt gcgtcaccag atggcccacc ggatatatat aaacagcata 840

aacaggaaat tggagcacat act 863

<210> 17
<211> 510
<212> DNA
<213> Homo sapien

<400> 17
gtcgcggcga ggtactaatt gtaaaattac cttgaccaaa ctgatatcag gatgacaatg 60

tcaacattgc acactctttg aagacagttc tcatttgctg aatgaatgaa tgaatgctga 120

cccacacatt ttcaagggaa aaccttggtt atgttcaata catgtatttt aaaactgagg 180

gaagtatgtc ctttctgaga gattgtcatc aacatggata actgaaaggt tttgatgctt 240

aaaaggataa gctccaaata tctggaacat tcacttgcat atatcactac atttcccaat 300

agtatttgtt aaattaaacc ttaggtaata aagtgctagg gttctttttc ctgttctaac 360

gaagtgaata tcataaatca attcattaga acaggaaaaa aaaaaaaaaa aaaaaaaaaa 420

aaaggctggg ggtaccctgg ggccaaagcg gtccccgggg ggaattggtt tcccgcccaa 480

attccccccа aaaaacaaaa aaaaggtgcc 510

<210> 18
<211> 947
<212> DNA
<213> Homo sapien

<400> 18
ccgcccgggc aggtacccag gacacaaaca ctgcggaagg ctgcagggac ctctgcctag 60

gaaagccagg tattgtccaa ggtttctccc catgtgatag tctgaaatat ggcctcgtgg 120

gaagggaaag acctgaccgt ccccccagcc cgacacccgt aaagggtctg tgctgaggag 180

gattagtgta agaagaagga acgcctcttt gcagttgaga caagaggaag gcatctgtct 240

ccttcccgtc cctgggcaat ggaatgtctc ggtataaaac ctgattgtat gttccatcta 300

cttaggggga aaccgcctta gggctggaga tgggacatgc gggcagcaat actgctcttt 360

aaggcattga gatgtttatg tgtatgcata tctaaaccac agcacttaat tctttacctt 420

gtctatgatg cagagacctt tgttcaggtg tttgtctgct gacctctcca caattatcct 480

atgaccctgc cacgtccccc tctccgagaa acacccaaga atgatcaata aatactaagg 540

gaactcagag gccggcggga tcctccatat gctgaacgct ggtccccttg ggccccctta 600

tttctttctc ttataagaca aaaaaaaaac aaaaaaaaag cctgggggta accctggccc 660

aagcgctggt ccccgtgggt tggaaaattt tttctgcgct cccatgctac cacaaaacaa 720

caacacaaca aaacaaaagt gtatgacaac aagagagcga gaaagaacaa acaaaaagaa 780

gagtaggaga acacgccata aacaagaata atgacaacat cacaagaaag caacaacacg 840

aacaacacat caaacactcg agaaaaagag aacaacaaag aaacacaaca gacacaacac 900

aagcactcaa accaaccagg acaacatcaa ctcacagata cacagga 947

<210> 19
<211> 854
<212> DNA
<213> Homo sapien

<400> 19
cggggtcgcg ggcgaggtcg ccaggaccga acgggaaggt tttccagttc tctgtaatgc 60

gacacatgcg agtcggtttt ccctttccat agtaggaagt cttcggacag tataagactg 120

gagctgtcct cttacagcat aggacggcac cataagggac atgtcggata gagatttcac 180

aagacgaaga actgccgcct gagctagttc agaatgtcca ggagggccag ttactaggtt 240

agcgagtcct tccttgtctg gcactcataa ttgggccaca gagggacact catgaatacc 300

atgtactgtg tagacaaggc atcagatcgc acgaagcgaa gctactcacg gtctcgcgta 360

acggcacagg ccatgagaac ggctacagca atgcagtcta taatagagac gtgcaatggg 420

cggtgatgct aatggcacct agcaagaggg catggaccgt ataggcaaca acgatttgtc 480

gaaagcctcg ttgagcatgt gttcgaggct ccctcgagtt tggttgtgaa aaacggctga 540

taactgtgca ggattttaga tttcttcaag atacatggga ttgggagtta tccagactac 600

caggaacaac aaaacaaaaa aaaaaaacaa agaaggctcg tgggggggac ccaaagggcc 660

gaaacgcggt gttccccggg ggtgggagaa agaggagaga cgcgggggcg agaaaaattc 720

tccccccaaa atcaggggcg gccacaacag acatatgtgg atacgagaaa acaaaagaaa 780

agaaaaaaga agaggagaaa caagaaataa aaaagaagaa agaaagaaag caaaaaaaca 840

gagaaaagaa aaaa 854

<210> 20
<211> 564
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (99)..(99)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (108)..(108)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (112)..(113)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (519)..(519)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (557)..(557)
<223> a, c, g or t

<400> 20
gtggtcgcgg cgaggtctcg tgagcccccct agaccatcac ggatgccgag cttcgggtaa 60

ctctcacagt ggaaggttcc cacgccgccc ctaatcccng cttcgganag gnncaagccc 120

tgagaaacat cgcccattct ctctccatat caccccccaa aaatttttgc caccccaaca 180

cttcaacact attttgtttt atttttctta ttaatataag acggcaggaa tgtcaggcct 240

ctgagcccaa gccaagccat cgcatcccct gtgacttgca cgtatatgcc cagatggcct 300

gaagtaactg aagaatcaca aaagaagtga atatgctctg ccccacctta actgatgacc 360

ttccaccaca aaagaagtgt aaatggccgg tccttgcttt aagtgatgac attaccttgt 420

gaaagtcctt ttcctggctc atcctggctc aaaaatcacc cccactgagc accttagcaa 480

ccccactcct gcctgccaga gaacaaaccc tctttgacnt gtattttcct ttacctaccc 540

aaatcctata aaacggnccc accc 564

<210> 21
<211> 5606
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (301)..(301)
<223> a, c, g or t

<400> 21
tttggtgctg tgactcggat tcggggacct cccttgggag atcaatcccc tgtactcctg 60

ttctttgctc cgtgagaaag atccacctat gatctcaggt cctcagaccg accagcccaa 120

ggaacatctc accaatttta aatccggagc ttgctacact tgctggaaat ctggccactg 180

ggcaaggaat gcccgcagcc cgggattcct cctaagccgc gtcccgtctg tgtgggaccc 240

cactgaaaat cggactgttc aactcacctg gcagccactc ccagagcccc tggaactctg 300

nccaaggctc tctgactgac tccttcccag atcttctcgg cttagcggct gaagactgac 360

actgcccgat cgcctcgtga gccccctaga ccatcacgga tgccgagctt cgggtaactc 420

tcacagtgga aggttcccac gccgccccta atcccgctcg aagcagccct gagaaacatc 480

gcccattctc tctccatatc accccccaaa aatttttgcc accccaacac ttcaacacta 540

ttttgtttta tttttcttat taatataaga cggcaggaat gtcaggcctc tgagcccaag 600

ccaagccatc gcatcccctg tgacttgcac gtatatgccc agatggcctg aagtaactga 660

agaatcacaa aagaagtgaa tatgctctgc cccaccttaa ctgatgacct tccaccacaa 720

aagaagtgta aatggccggt ccttgcttta agtgatgaca ttaccttgtg aaagtccttt 780

tcctggctca tcctggctca aaaatcaccc ccactgagca ccttgcaacc cccactcctg 840

cctgccagag aacaaaccct ctttgactgt aattttcctt tacctaccca aatcctataa 900

aacggcccca ccccttatctc ccttcgctga ctctcttttc ggactcagcc cgcctgcacc 960

caggtgaaat aaacagcttt atggctcaca caaagcctgt ttggtggtct cttcacacgg 1020

acgtgcatga aatttggtgc cgtgactcga atcgggggac ctcccttggg agatcaatcc 1080

cccgtcctcc tgctctttgc tccctgagaa agatccacct acgacctcag gtcctcagac 1140

cgaccagccc gagaaacatc tcaccaattt caaatccgag acaaaggaga cacattttat 1200

ccgtggaccc aaaactctgg tgccggtcac ggactcagga aggcagcctt cccttggtgt 1260

ttaattattg cagggacacc tctctgatta tttacccacg tttcaaaggt gtcagaccac 1320
gcagggatgc ctgccttggt ctttcgccct tagtggcaag tcccgctttt ctggggaagg 1380
ggcaaaccat cacggacgcc gagcttcagg tgactctcac agtggaagaa atatatctct 1440
gtgctgccta ttatccaatg tctgaaaact gttgtttcat atatattttg cccagatttt 1500
taggtaagga agtagtaaga gaaagacaag aaaacctaag acttgtgaga ctgatgcaag 1560
ataaagagga aatgattgga aaactcaaag aagaaattga tttattaaat agagtaaata 1620
atgaaattaa gacagaaatc aagaagttct ttgaaaccag tgagaacaaa gatacaacat 1680
accagaatct ctgggacgca ttcaaagcag tgtgtagagg gaaatttata gcactaaatg 1740
cccacaagag aaagcaggaa agatccaaaa ttgacatcct aacatcacaa ttaaaagaac 1800
tagaaaagca agagcaaacg cattcaaaag ctagcagaag gcaagaaata actgaaatca 1860
gagcagaact gaaggaaata gagacacaaa aaacccttca aaaaattaat gaatccagga 1920
gctggttttt tgaaaggatc aacaaaattg atagaccgct agcaagacta ataaagaaaa 1980
aaagacagaa gaatcaaata gatgcaataa aaaatgataa aggggatatc accaccgatc 2040
ccacagaaat acaaactacc atcagagaat actacaaaca cctctacgca aataaactag 2100
aaaatctaga agaaatggat aaattcctgg acacatacac tctcccaaga ctaaaccagg 2160
aagaagctga atctctgaat agaccaataa caggatctga aattgtggca ataatcaata 2220
gcttaccaac caaaaagagt ccaggaccag atggattcac agccgaattc taccagaggt 2280
acaaggagga actggtacca ttccttctga aactattcca atcaatagaa aaagagggaa 2340
tcctccctaa ctcattttat gaggccagca tcattctgat accaaagctg ggcagagaca 2400
caaccaaaaa agagaatttt agaccaatat ccttgatgaa cactgatgca aaaatcctca 2460
ataaaatact gacaaaccga atccagcagc acatcaaaaa gcttatccac catgatcaag 2520
tgggcttcat ccctgggatg caaggctggt tcaatatatg caaatcaata aatgtaatcc 2580
agtatataaa cagagccaaa gacaaaaacc acatgattat ctcaatagat gcagaaaaag 2640
cctttgacaa aattcaacaa cccttcatgc taaaaactct caataaatta ggtattgatg 2700
ggacgtattt caaaataata agagctatct atgacaaacc cacagccaat atcatactga 2760
atgggcaaaa actggaagca ttccctttga aaactggcac aagacaggga tgccctctct 2820
caccagtcct attcaatgta gtgttggaag ttctggccag ggcaattagg caggagaagg 2880
aaataaaggg tattcaaata ggaaaagagg aagtcaaatt gtccctgttt gcagatgaca 2940
tgattgtata tctagaaaac cccattgtct cagcccaaaa tctccttaag ctgataagca 3000

acttcagcaa agtctcaggc tacaaaatca atgtgcaaaa atcacaaaga ataaaatacc 3060
taggaatcca acttacaagg gatgtgaagg acctcttcaa gaagaactac aaaccgctgc 3120
tcaaggaaat aaaagaggat acaaacaaat ggaagaacat tccatgctca tggataggaa 3180
gaatcaatat catgaaaatg gccatactgc ccagggtaat ttacagattc aatgccatcc 3240
ccatcaagct accaatgcct ttcttcacag aattggaaaa aactacttta aagttcatat 3300
ggaacgaaaa aacagcccgc atcgccaagt taatcctaag ccaaaagaac aaagctggag 3360
gcatcacatt acctgacttc aaactatact acaagcctac agtaaccaaa acagcatggt 3420
actggtacca aaacagagat atagatcaat ggaacagaac agagccctca gaaataacgc 3480
cgcataccta caactatcgg atctttgaca aacctgagaa aaacaagcaa tggggaaagg 3540
attccctatt taataaatgg tgctgggaaa actggctagc catatgtaga aagctgaaac 3600
tggatccctt ccttacacct tctacaaaaa tcaattcaag atggattaaa gatttaaatg 3660
ttagacctaa aaccataaaa accctagaag aaaacctagg cattaccatt caggacatag 3720
gcatgggcaa ggacttcatg tctaaaacac caaaagcaat ggcaacaaaa gccaaaattg 3780
acaaatggga tctaattaaa ctaaagagct tctgcacagc aaaagaaact accatcagag 3840
tgaacaggca acctacaaaa tgggagaaaa tttttgcaac ctactcatct gacaaagggc 3900
taatatccag aatctacaat gaactcaaac acatttacaa gaaaaaaaca aacagcccca 3960
tcaaaaagtg gatgaaggac atgaacagac acttctcaaa agaagacatt tatgcagcca 4020
aaaaacacat gaaaaaatgc tcatcatcac tggccatcag agaaatgcaa atcaaaacca 4080
caatgagata ccatctcaca ccagttagaa tggcaatcat taaaaagtca ggaagcaaca 4140
gaaaccttaa caaggctagc aacagaggtg gcagtcaagt ggtccaggaa gggcgggaag 4200
aaaacaggag gcaagaaaac accagacggg gcaccgaaaa ccgggaaaac gccgacgaca 4260
gggaagaagg agaacaacgc ggcaccaaca cccggcacaa aacggccaca aaccacaagc 4320
gggggaagcg ggaccgagac cacagacgcg gaaagaggga aaaaaggaga aaacggaaaa 4380
gcggcgggca gcccgaggag agggggaaaa gagggcggag aagaggggac aggttgaaga 4440
cagcggaccg acggcgggcc cccgacacaa agacccgcac ccgcaccagc caccagcccg 4500
ccccccctagg gcccagccca catcaacgta acgcacctct tacccaaagc tcggcgtccg 4560
tgatggtcca gggggcttcc gaggcgatcg ggcagcatga gtcttcagcc gctaagccga 4620
gaagatctgg gaaggagtca gtcagagagc cttgggccag agttccaggg gctctgggag 4680
tggctgccag cgcacctgcg ccctgtctgc gccgccagtt tcctgaacac aggcagcagt 4740
catgttctgc tcacccacca cccccagcgc tgtggttcct catcacagac accggagtca 4800

caccgtggct ctcaagtcac tgtttccaga acttcgcctg tgcatacaat aagctgccac 4860

agagaagcaa taccaccatg gcatggcgtt gggccggccc ttactgtact gcaacaggca 4920

acttctcaca tgggggagtt aagccatttg agtgcttaac atgtggagta gcttgggctg 4980

atgcccgatc tctaaaacgc catgtcagaa cacatactgg tgaacggccc tatgtctgtc 5040

ctgtatgtag cgaagcctac atagatgctc gaacactccg ccctgaaatc ctgaagagat 5100

taatcgagat tctagcccct tttaaagatg tttcagaaat gttggaagaa ttgggactag 5160

accttgaatc tggagaagaa agtaatgaat cggaagatga cctgagcaac tttacttcat 5220

ctccaactac agcatccaag cctgcaaaaa agaagccagt atccaaacat gaacttgtgt 5280

ttgttgacac caaaggagtg gtaaaacgtt cttctccaaa acattgtcag gctgtcttaa 5340

aacagctgaa cgaacagaga ctttccaacc agttctgtga tgttactttg ttaattgaag 5400

gagaagagta caaagctcat aaatctgttt tgtcagcaaa tagcgagtat ttcgagtctt 5460

tttatgagaa aggagctgta accagctggc tgcatctcag gccttacttt agaccaagct 5520

tgtccaaccc gcggcccatg ggccggccac atgcggccca ggacggcttt gaatgtggcc 5580

cagcacaaat ttgtaaactt tcttaa 5606

<210> 22
<211> 797
<212> DNA
<213> Homo sapien

<400> 22
gccgcccggg caggtacaaa aagaactccc acaactcaac aacaaaaaaa ctgttcaaaa 60

atgggcaaag gtttgaatgt ggtttttgtt gcagtctatc ttaaaaaaaa aaaaaacagg 120

caaggacttg aatagactcc ttcttaagag gatatgcaat tggccaataa gcttctaacc 180

agaattctgg ttttcactaa tcattaagga aatgctaatc aaaaccacaa gataccacct 240

catgcccatt acaatggcta ctatcaaaag aacaaaaaat aagtttggaa gaatgtggac 300

tcaattggga actcttgtat actgtcggtg ggaaagtgaa atgtgtatag ccactgtgga 360

caatgttcag tcctcaaaaa attaaagcta gaattggcca aatgaaccag caattccact 420

tctgggtata tgcccaaaga aaacagggct gggcgcagtg gcttacactt gtgatcccag 480

cactttggga ggccaagttg gggcggatca caaggttcgg agatcgagac catcctggcc 540

aacatggtga aaccctgtct ctactaagat tgaaaaaatt agctgggcgt ggtggctggc 600

acctgtaatc ccggcccgtt tgacagtcaa ggccgggtgg ataacctgag gtcagagagt 660

tccagaccag cttggccaat gttggcaaaa ccctggtctc tcttagatcc aaattaggtt 720

ggctggtggg cttgcccttt tatcccgcta tccgggagct gggcggaatt tggttgactt 780

taagcgcaag gttgcta 797

<210> 23
<211> 666
<212> DNA
<213> Homo sapien

<400> 23
gcggccgccc gggcaggtac ccaacagctc attgagaacg ggccaggatg acgatggcgg 60

ctttgtggaa tagaaaggcg ggaaacggtg gggaaaagat tgagaaatcg gatggttgcc 120

gtgtctgtgt ggaaatgaag tagacatggg agacttttca ttttgttcta cactaagaaa 180

aattcctctg ccttgggatc ctgttgatct gtgaccttac ccctcaaccc tgtgctctct 240

gaaacatgtg cgtgtgtcca ctcagggtta aatggattaa cgggcaggtg caatgatgtg 300

ctttgttaaa gcagatgctt gagggcagca tgctcgctta agagtcatca ccaaatccct 360

aatctcatag taatcaggga cacaaacact gtcggaaggc cgcagggctc ctctgcctag 420

tgaaaaccag agaccctttg ttcacttgtt tatctgctga ccgttccgct ccactattgt 480

cccatgaccc gctgccaatg ttatcccctc tgtgaagaag acacccaatg aacttatcta 540

ataggaaaaa aaaaaaaaaa aaaagggttg gggtggaccc tgggccaaag cggttcccgg 600

gggggaaatt tggttttccg ggccacaatt tccccaatta aaaaaaaaaa ggcaaaagag 660

gaaaaa 666

<210> 24
<211> 975
<212> DNA
<213> Homo sapien

<400> 24
aatgttcgat cactataggc gattggttat ctagatcatg ctcgagccgg cgcgttatgt 60

gaatggatag cttccggccg gagggtgggt gacagatcgt cacccatgga tgcacacaag 120

cgagctggta aggagagcat gggagtccgg gacagagagc tcaccatatg cgcctcagaa 180

acgttcgcgg atgctagcat ctggcaagga ctagacttcg gaatacttgt gagcgggaat 240

gtggtaccat gtgttgaaca cccaggagtt tgtcctcttc ctgttcgccc ttttctctga 300

tttaccattg gacagtccct tagagtgctt ggccaatgga tccgtcgtgg atgtccgata 360

cgccttcctg agtcgtcggt cgtacttgga ggtcgtccat ccaaggtggg tacggcgtgt 420

tgtggtgtcc cttgattgtt ttctggtcgt tctcttagga ggaagccccc ctccactttg 480

gtgtcctccc tggtgagcca gcctcgcgac ctgccgagag caccccgtcg tggtcctcgg 540

tatagtgagt aatggggctc agggcctctc caacaacaga gaggagctga tgctgtaggg 600

ctgaccccgt gacttcctga gtcctcaccc tgtccagtgc tttgagattc ttcccacctc 660

cccatcctca ccagccggat cgggcgctgt gcagtgtggt cagcattggg tgaagaaagt 720

catttcctcg ttggggcagg tattcctctt tatctctcat tacactggaa atgtttattt 780

ctgctgtatc atccgtgctc aaacgtttaa gttctgtcag gctcaccttc tctctggaaa 840

gaatttgctt aacttgacat tccatggtgc ccgctaataa aatatatttt gaaccaaaaa 900

aaaaaaaaaa aaaacgctgg ggtacccggg caaaaccgtc ccggtgaaat ggtcccccac 960

accaaaaaaa aaagg 975

<210> 25
<211> 554
<212> DNA
<213> Homo sapien

<400> 25
tgccgaccgc gcgtagtgat ggatgagcgg cgcccgggca ggtacctggg gcttccagaa 60

tataccaaat tcagagagcc cacattcacc cctgactgtg cttggagcaa acctgaaagt 120

tcactcccaa gaggcttgtt ccagcccatt cctttattct ggaaagtaat tcttggtatt 180

gaaacggaaa actgggacaa agggagcttg aggaaaacga aaacaaacaa tgaaactgga 240

gatatgctat ttagtcttaa cccatcacag atatgctgtt tggccttaac ccatgtggaa 300

atttgtaaac tttgtcagga tttccctgtg catggtggtg aaagtcatgt agggaagaaa 360

aaattcaccg tgtaatctct ttagaagtca caaaaaaaaa aaaagctggg cgttaatcag 420

ggccataggc tgttcccggt ggtgaaatgg ttatccgcct ccacaattcc cacacaacaa 480

gggaagccaa ggtaaggcag gaggagaaag agagagcgag ggaaaagaca aaccaaacaa 540

caaccaaaaa cgcc 554

<210> 26
<211> 2581
<212> DNA
<213> Homo sapien

<400> 26
ccccaagaag cattcagtag ccttgtgctc ctaacttgtg accttgtgtg tggcgtggct 60

cccccactaa gtgtaaattt gtgttttcaa gccttccgag gacgaagtgg taagatgaaa 120

gctggcgttg ccttcgtgtt tatccgggct tctctcttcc catccttgtg atcactctgc 180

tgaccctgcc acaccctggg gccatacaca tgactcctgt gcctctgcat ctcactggcc 240

atttcaaacc agttcgtgag cctgtgttct aaaaagctca tctccattaa ctgcatctca 300

tgtcggtcac ctctgtcttt gtttcattga ctttgctgat ttaacacctc tgagaatggc 360

tttgaaagga gctacaagta atgtcttcac tggcattctt gaagtgactc ctcttggggg 420

ttaagtccca ttggcctctg ccttcctctc ctgtgatgtt gtggttgatc tactcccagc 480

caccgggagt agccgtttcc tccctgctcc tcctagtgct tgtgagcaaa ccttaccctt 540

cctccctgtc tgagccagcc tcttcctgct gccctcctca gcttgatgat gcttcaactt 600

agaagtggtg gacccttctc ggggtcagct ctacctccat gtgacccagg tgagagggag 660

acagctttca aaaggaggct ttgccttcca gatgcatccc aaaggaaata atccattagt 720

agaggcattt tgtggaggag tgaaggtgga agccagggtg cagtagggtg aggactggct 780

gggcagagag gaaactgagg cagaagatgt ggacagggac tctgcagcga ttcatggttg 840

gtcatggccg attgatacgg gcgtgtgctg agtggcagac tcataagtgc tgagctggat 900

ttctcaccca gtccttgctt ctcctgagaa cacctgcagg aagctgggaa tgaaggagag 960

aaggcaaggg taaggcctgg aagcgaggct tctctcctgt cctgcagttt attcggtggg 1020

agaaacttga ggtagaataa aaacactgaa gggattggct gtaagggagc ccagtggtca 1080

ctgccctccc agcccatagg agtagagatc agcactcctg ggagccatgg acttcttcct 1140

ctcaatgcgg gccctggggc ggcagccagc agtggtagga aagagctctt aaccacggtg 1200

ttgggattta gggaactcag ggaactagga ttaccctagt tttcagggta agtaaggaag 1260

cttgaagagc ctgttccgga agaaaagccc agaaataggc taggtgtcga cagaggtgag 1320

ggagaagcag acgttcgccc agagatggca tggtggagcc agccccctcc acgctggctg 1380

cacccctgtg gcctgggctg ttgtggcttt accagccaag ccactttctc ccatccccat 1440

gttctcagtg atttcccaag gcatttctgg tgtcctttgg gagtggaatc acgtgggttt 1500

tgaaaggaac tgcggtgaaa aatagaccct gacctgaggg caggggccga tggggaggcc 1560

actgagccct caggagtgtc tgaaagctca gaagattctg gtctgggctc tgcgggggga 1620

agccccataa gggagcgcct cctggtcttt ggtcagcttg acagagaggc ccagggaaag 1680

tacctggggc ttccagaata taccaaattc agaagcccac attcacccct gactgtgctt 1740

ggagcaaacc tgaaagttca ctcccaagag cttgttccag cccattcctt tattctggaa 1800

ataattcttg gtattgaaag gaaaactggg acaaagggag cttgaggaaa acgaaaacaa 1860

acaagaaact ggagatatgc tatttagtct taacccaaac agatatgctg tttggcctta 1920

acccatgtgg aaatttgtaa actttgtcag gatttccctg tgcatggtgg tgaaatcatg 1980

tagggaaaaa aaaattcacc gtgtaatctc tttagaagtc acaaaagaaa aagaaggtta 2040

tcttgctccc aaaaggctgt aaaaagaata agtaaagtgg ccatagaggc ctagtcttct 2100

caggacagtg tccgggttga gagtctgtct cctgaagcgc actctgggga aaatcccttc 2160

ctgccctcct gcaggtcctt agggtcccag acccagacag tcactttctc aacagagtgc 2220

cgtcaactca gcacacactc ctctcttgag cacagagccc cagagggaga agaacaaatg 2280

tgttgaaaag aatcttatta agatgtagtt aattaaaatg taatgtattg aggggaatgg 2340

aggtgtccca ggtgagggct aagtcaggca ggatttttgg ggaaggcatt gccgaaatca 2400

ccacctgagc tcaacactgg gtgcttctgg cccctccaga gttgaggtgc catccatggg 2460

aagtgcagtc ccctgcctgg cccaggttca aagcgccaag tagccacaac tcagaatgcc 2520

tgcacgttcc cctcctagcc ttatatcttc tctctggttt cctcccacga cagtttgaca 2580

t 2581

<210> 27
<211> 657
<212> DNA
<213> Homo sapien

<400> 27
gggtgctata agcatggtct taatcagctc cgaccggcgc agttgtgatg gattggtcgc 60

ggcgaggtat tgttagcatt cccattttac agtggaggaa gctgaggctc agagatgtta 120

agcaagctta gctgaatggc cacaccacca gcgaagtgcc tgagccaaga tttggactcg 180

agtccatggg acccccacgc tcgtgaggct gactgctctg ctcccactgg gtcccttcat 240

gaggtcgtcc cacagcactg ctagttccag ggcgagtgcc agcacatggc cccactggga 300

gccgggggcc tgatttaggt ctactggaaa aagtgtcacc tttggggaca ctcaaggcac 360

aggctggttg gtttcgttgc tggattttat atactcatgc cctaaccctg tgttcctggt 420

ttctataagg ccccggggca aggtgcaagg aatttgcaaa tagggcctgt atgacttatt 480

tcctaggaca cgggaagctt ttcttacctc ctttctaccc tcttctccaa cctgaactcc 540

caagtttctt ctcctgaagg tctttgcact ataagcgcca aggagcccgt gtgcgtggca 600

ggggcggctg ggagggtatc tggagaacct tagtgaggcc tctggcctag ccagaga 657

<210> 28
<211> 1244
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature

<222> (37)..(37)
<223> a, c, g or t

<400> 28
tgactggagt tccatgaggg agggaaattg atgtcanagt gtcattttaa agcttaagct 60

gaaagtttat tttttaaatt ctcattcatt catttagcat atattgattg agcatctaca 120

atgtgccagt tgtagaattc catctcagaa gagacttgac ttgtggatgg tggaggggca 180

gtcctgctcg gaagcagatg atgtgaaatg ttcctttcca gtctggttca cgatgtttaa 240

cagatttgtc aggtcaccac tgtgacccca agctttgctg gcagattgtt atatagtatt 300

tactgagagc cctgctatct ggtaaaggca gttaaaaagc ctgcaatctc gactcatttc 360

cagcatgaac agactggtcc ttgctgcttt acacaataat caaagctacc ttttatggcg 420

tgctcgccac tcccaagcac tgggcgaagt gctttacccg tcttcccctc cgcgatgcct 480

catgcccact ttagcagata gtactgttag cattcccatt ttacagtgga ggaagctgag 540

gctcagagag gttaagcaag cttagctgaa tggccacacc accagcgaag tgcctgagcc 600

aagatttgga ctcgagtcca tgggaccccc acgctcgtga gctgactgct ctgctccact 660

gggtcccttc atgaggtcgt cccacagcac tgctagttcc agggcgagtg ccagcacatg 720

gccccactgg gagccggggc ctgacttagg tctactggaa aaagtgtcac ctttggggac 780

actcaaggac aggctggttg gtttcgttgt ggattttata tactcatgcc ctaaccctgt 840

gttcctggtt tctataaggc cccggggaag gtgcaaggaa tttgcaaata gggcctgtat 900

gacttatttc ctaggacacg ggaagctttt cttacctcct ttctaccctc ttctccaacc 960

tgaactccca agtttcttct cctgaaggtc tttgcactat aagcgccaag gagcccgtgt 1020

gcgtggcagg ggcggctggg agggtatctg gagaacctta gtgaggcctc tggcctagcc 1080

agagaggcaa taagcttggg gacgttccgt tctgggttct gacgttgttg gttctgacgt 1140

cgttgtgctc ttttgtaaga ggaatttcat accttggaga cgctttgtac atatttgtaa 1200

tgactttatt aaaaaactga ttgtgcactt ctaaaaaaaa aaaa 1244

<210> 29
<211> 663
<212> DNA
<213> Homo sapien

<400> 29
tcctagtatg catgctcgag ccgcgcgtat gtgatggatg tcgcggcgag gtaccgaaag 60

tgagcggggc aggcacgcta gtcacatggg taatgtggca gggtgtcgtg tcactgtgct 120

ttggctccag ggccagagca gtctgactta gtgttgagct ccaagcatgg aacacttgga 180

gtttggttca tttttgacca gcaagcctct aaatgtggtg ccttgattac ccaccgcaag 240

ggagagtggc agttgccttt ttatgacatg ttaattccag ccaggtgagt caccaggtag 300

ctctcatcct cctgccaggc tcccgctgcc tgtcggtttg gcattgtcag actagatggt 360

gactcagtgt cattggaagg tgacagtttg aggttccaaa ccagttttct cctttaacca 420

tttcaccctc aggagtgatt cctcctttgt ttggcattgt cagggaatgt gatgatccat 480

tcaaatgact tttggagttc caaatagtgt ttctacttta acttccaaaa aaaaaaaaag 540

aaaaaaaaaa aaagggcggg gggtaccctg ggccaggcgt gtcccggtgg tggaattgtt 600

tttcccgtcc caattccccc catttttcac aacaatggtg agcctggtca aaagagaaaa 660

act 663

<210> 30
<211> 643
<212> DNA
<213> Homo sapien

<400> 30
cagccgcccc acacaccccc gatgatcgat cactatgggc gaattgtgct ctagaatgct 60

gctcgagcgg cgccattgtg atggatgcgt ggtcgcggcg aggctttgtg ttaagcgtga 120

ggcagaggga gacgttagtc cagacatttc caaagtgtgg gtgggtccgt tggttcccga 180

gatactttta ggtggtatgg ggcctgcatt aagtggcaca aaaatcagag caagaaagcg 240

atgcccttcc ccaattctct caatcctttt tatggccgag aagatctcag ctggatgcca 300

acatgttccg atgcctgtgg aagacatgcc gacgtctcct ctgcctaggg agcaggactt 360

gggcttaggg caggtggaaa aaattccaga ctttttttaga cactgtattt tgttttaatg 420

gtatattcta tattggctac tttattgtat aggacaagtg gtagtggcat tctatttatt 480

ggtgaccttt tcaataaata gatttaagca aaaaaaacaa aaaaaaaaaa aaaaaagctg 540

tgggggttac ccggggccaa aggggggccc cggggggggaa tgtggttctc ccgccccaat 600

tccccccaaa tttttgacaa aatgaagagg acacacggaa ccc 643

<210> 31
<211> 1192
<212> DNA
<213> Homo sapien

<400> 31
aaaccgcctg gagccgccgg gagtggacgc cgccgaggcc cggagtcgcg cctgcagaca 60

cagcatctac tcagcgtggg tcacctctgt gaacatcact gactgcaagc ctccctcaat 120

ttctggtgca gcccatcagg gacccacagc gcctgggagg atggtgcgga tcttggccaa 180

tggggaaatc gtgcaggacg acgacccccg agtgaggacc actacccagc caccaagagg 240

tagcattcct cgacagagct tcttcaatag gggccatggt gctcccccag ggggtcctgg 300

cccccgccag cagcaggcag gtgccaggct gggtgctgct cagtcccccct tcaatgacct 360

caaccggcag ctggtgaaca tgggctttcc gcagtggcat ctcggcaacc atgctgtgga 420

gccggtgacc tccatcctgc tcctcttcct gctcatgatg cttggtgttc gtggcctcct 480

cctggttggc cttgtctacc tggtgtccca cctgagtcag cggtgacctc tgagggctga 540

taggggtggg tttgttgaga gggacttgct gggccttggt gtgagagcag gcatatttgg 600

aggggatctg gtggtgcctt gaaggtatga tcagagaggg gaccacaggt gtgtgtttcc 660

cctttgtgtt aagcgtgagg cagagggaga cgttagtcca gcatttccaa agtgtgggtg 720

ggtccgttgg ttcccgagat acttttaggt ggtatggggc ctgcattaag tggcacaaaa 780

tcagagcaag aaagcgatgc ccttcccaat tctctcaatc cttttatgcc gagaagatct 840

cagctggatg ccaacatgtt ccgatgcctg tggaagacat gccgacgtct cctctgccta 900

gggagcagga cttgggctta gggcaggtgg aaaaaattcc agactttttt agcactcttt 960

ttgttttaat ggtatatttt tattggctac tttattgttt aggacaagtg gtagtggcat 1020

tctatttatt gtgacctttt caataaatag atttaagtaa aaaaaaaaaa aaaaaaaaaa 1080

aaaaagctgt gggggttacc cggggccaaa ggggggcccc ggggggggaat gtggttctcc 1140

cgccccaatt ccccccaaat ttttgacaaa atgaagagga cacacggaac cc 1192

<210> 32
<211> 582
<212> DNA
<213> Homo sapien

<400> 32
aagtaaaaaa acaccacgag acaggtatga tatagactca tatggcgatg gtcctctaat 60

catctcgagc ggcgacagtg tgatggatcc tgcccgggca ggtacgcggg gggggttcc 120

tgcccccccc gcgcatggtg gaggtaggct cggaccggcc cgcggagctt gctgcagtcc 180

ttcgcgccct cctcgccctc cccaccgaca tcatgctcca gttcctgctt ggatttacac 240

gtgggcaagc gtgtgttgga agtgtactct ggctcagaca ctatgatata ccaaacctgg 300

gctaaaaaca ctttgagaga acaattacaa agaaggactt ggaatgccaa gaaaggaacc 360

ccccctagtt gccatgagac gtgcctccag caacttgccc ttcagcgata tacgtgattc 420

tacgtggctc ttgagggcct cgttaactat ctgaaccaaa gagcttggcg taatcagtgg 480

tcatagctgt tctcctggtg tgaacacttt gctataactc ccgggcttcc aacaaaaatt 540

ttccccaaca acaaaacata aggaaaaaca caaaaaaggg gt 582

<210> 33
<211> 900
<212> DNA
<213> Homo sapien

<400> 33
gggggggagg aggatgaaga actcactatg ggcgaatggg cctctagatg ctgctcgagc 60

ggcgcagtgt gaatggattc gcggccgagg tacactggcg aatattctta tttctgcaag 120

tttgcttaga ggttggcaac tgaagctgtg caggacgatt cctgttctgt aagattagtc 180

tccagttgtc agtcaagcag ttgagtgcgg tatgtctagt gcccagtttc cctctccaca 240

ggtccccata ggctcttctt gttaacttta caatccgcga tcagagatga gatctctgcc 300

aaggcagcaa ctgcaaggac catgtgggtc aatgttacca gcagacactc aaagcccatt 360

cccatttact tcaagcaccg cttttatagg attatcgttg agagacgtgg gtcatggttg 420

gtattatgag gtgagtggtc gagtgacatt cacgatttct cgatctttct gaatgcatag 480

tggctgggag tggtggctca tgcctgtgat cccggcagtt tgcggagggc cgcaggtgga 540

cagattgttt gacgcacagg cagttcgaga ccagccgggg taacatgggc gggaccccca 600

atctctacca aaaaaaaaaa aaaaatacaa caagtggcct gggtgtggtg tgcatgcctg 660

tagtcccaag tcccagctac tctaactgtg ggaggcgttg aggcagaagg atcacctgag 720

cccagggaag ggccagggct ggcgtggccc tgtatggtgc ccctgcactt gacctgtggc 780

aacgatgaga tgagccctgt caaaaaaaaa aaaaaaaaaa cgttggggat ccggccaagc 840

ggccggtgaa tggtcccgcc atcacccaac aaaaaaaagg aaaaaaaaaa aaacaaaaac 900

<210> 34
<211> 548
<212> DNA
<213> Homo sapien

<400> 34
cggccggccc gggtagggac tccagtcctg cgcgacatgt gagactcctt ctctgatagc 60

tatggaagat gtcgtatcta gaacatgatg atccagggtt ttgtctcatt cagatcatca 120

aattgtttat atagaagagt tggccttgac atgtgtgctc atgtctgcct tgtcctataa 180

gtaactaggc ttaagatgtc tttctaggtg tgtctcctgc atgagctaag gcattccttg 240

cacaagttag actgatttat gaaagtactg aaggacgcct ggtgatatgc caccacaaag 300

atatatccca ttgagcagct gtgggtgagt actgaggagc gtgagcttgt ccctggagca 360

tgatgctttt accgaagttc atgttaccct gctaaaagct gttccctcca atttgtgact 420

atggacctat ggagttggga caatctctat gggaagcaga aggcaaggac cccggtcatt 480

ttaggtagaa acaacagcat gctaatgcaa aaaattatgc agtgtgctac tgaacttcag 540

aggtgatc 548

<210> 35
<211> 372
<212> DNA
<213> Homo sapien

<400> 35
ccgcccgggc aggttgagcc cgtgcctgtc ttcagcatcc tcacagcaca gtgcctggga 60

gggtggagcc accagcctct ccctgaataa ctgggagatg aaacaggaag ctctatgaca 120

cacttgatcg aatatgacag acactgaaaa tcacgactca tccccctcca gcacctctac 180

ctgttgcccg ccgatcacag ccggaatgca gctgaaagat tccctggggc ctggttccaa 240

ccgcccactg tggactctga ggcctctgca tttgcgggtg gtctgcctgt gatattttgg 300

tcatgggctg gtctggtcgg tttcccattt gtctggccag tctctgtgtg tcttaatccc 360

ttgtccttca tt 372

<210> 36
<211> 734
<212> DNA
<213> Homo sapien

<400> 36
tcccttctta agtcgtgtgt acatcattgg gaatggaggg aaataaatga ctggatggtc 60

gctgcttttt aagtttcaaa ttgacattcc agacaagcgg tgcctgagcc cgtgcctgtc 120

ttcagatctt cacagcacag ttcctgggaa ggtggagcca ccagcctctc cctgaataac 180

tgggagatga aacaggaagc tctatgacac acttgatcga atatgacaga cactgaaaat 240

cacgactcat ccccctccag cacctctacc tgttgcccgc cgatcacagc cggaatgcag 300

ctgaaagatt ccctggggcc tggttccaac cgcccactgt ggactctgag gcctctgcat 360

ttgcgggtgg tctgcctgtg atattttggt catgggctgg tctggtcggt ttcccatttg 420

tctggccagt ctctgtgtgt cttaatccct tgtccttcat taaaagcaaa actaaagaaa 480

atacaccacc cccacaacca ccatcttgtc gacacgccct cggacttaaa aaatcgggcc 540

ctcgagaaaa gctttctaaa accttccgtt ggggcgcccg gggccccagt aggaagtcga 600

ccctggcaac agctgtcccg gaagaccggc acgacaaggg ctggttcccc cataaaaaag 660

cggcgaaacg gggttcgaca atccgaagga ttcgggccta cggtcttcac cgggccacga 720

ccaatatccc ggcg 734

<210> 37
<211> 537
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (492)..(492)
<223> a, c, g or t

<400> 37
gtttcctcag acagtcttcc ttgagcaact tcctaaacgc ccttcattat cccctttcaa 60

gctcatgcct aggagcaagg agcaaggcca ttagaaaggc ttcatccacc agcaggagaa 120

gctaggacat cccaaagggt ccacttctag agaggtgcca acccccacac gcacaccagg 180

cacacaaatg catgtgtgca cacgcatacc acaccctcca attgtcccca gaatggctcc 240

cttcagggag tcatgttccg ggacaccaaa tgagggaaaa tatccttctc ctacagtttc 300

ctgctcacat ttggattgag aaatgggatg tctctaaata atgtgtctaa attctcttaa 360

cataagtgca tatgttacgt gaaaaaaaca aaaaaaaaaa aaaaaaaggt cggggggaac 420

ccggggccaa ggcggtcccc gggggaatt tggttccccg ccccaattcc cccccattcg 480

ggagaaacag gntggcggcc gagaaaaccc gggcaccaag aaagccggac acacacc 537

<210> 38
<211> 1778
<212> DNA
<213> Homo sapien

<400> 38
ggccgcctca gtttccctgg ccgtaaaacg cgggcgatgg cgctcgatgc cccgggagac 60

gccggggtcc gcgggcccctt ccggcaccgg gtcctccttc aggttcgggc acccggctgc 120

ctccacgcgg ggcgagcgcg gcgatgggat cggtccgtgc tgcgctcagc cccgctttcg 180

ctatccctct gctccagggg cagcaaaatc agcagatgag cccccagcgc gcctgcggcc 240

cacacactgc ggccgtaaca gttaaacaga cggggaacat tcatgtgcgt agagctcttc 300

acggggcgat acagaaatgt gtcgttctaa agcgttaacg acctgcaatt tgcatatcct 360

ccaacacatg gattttactg attcaattat ggatgtgtgc aaagaatctg ctgctcgggc 420

gtggtgtcct cagtgctgag agagccccgc gacggcggtg cagctcctcc agggcgcgca 480

gccccgccac ctggcgccac cgcgagggag cgcaggccca gggcccgggc agacgctgga 540

```
ggggacgcgg caggtgaaga tggggcacct ccacttccct tctaaaggag cctgggaata    600
ggctaggaaa tgtccccact gtagaagaca gcaggaacat ctgacacgcg gataacctgt    660
gctgagctcg ttatattccc agcatttaat tctagccatt aggcatgcgt tagccacaac    720
ttcgcatagg aaggttattg caggtaagtt aagactcctt tccacttctt cctaagggtc    780
gctgacccat ggttctctag cttttgctgg taatcccacc cagattctca gcaccctagg    840
ttggaatgac ccctcctgag gtgggtgggg tagctgggag gcacctccct gaactgactt    900
ttctgccttt ttgaaggtgc agtggctgag cgctccttca gtgttttccg gtagtctttc    960
ttcagcccag tacgcagttt cctcagacag tcttccttga gcaacttcct aaacgcccct   1020
cattatcccc cttcaagctc atgcctagga gcaaggagca aagccattag aaaggcttca   1080
tccaccagca ggagaagcta ggacatccca aagggtccac ttctagagag gtgccaaccc   1140
ccacacgcac accaggcaca caaatgcatg tgtgcacacg cataccacac cctccaattg   1200
tccccagaat ggctcccttc agggagtcat gttccgggac accaaatgag ggaaaatatc   1260
cttctcctac agtttcctgc tcacatttgg attgagaaat gggatgtctc taaataatgt   1320
gtctaaattc tcttaacata agtgcatatg ttacgagaga tctgatccca gcctcccttt   1380
tttgttaaag tggtggtgct ttgccaccca tgtcaaaatg atctgggctt tccacatcaa   1440
atgcaaggga tgagcttgga ggacataccc cagtaccagc aatgatcttg agccaatggc   1500
agcgtcagct cacaacgggg tgattgaggc ttcttggtta gaagctttag aaacttgagg   1560
tcagtaaaac ccaaattgat cctttctcta gatcctctag atttctctag aggaaaaggt   1620
gacaaaaaca gacacttgtg ctcaccgtca ggggcacaat gctgtccgtc aatcctcatg   1680
cagtcctccc caccatttcg cacactgaca ccatcactac tttatgataa aaggagactc   1740
aggcagctta gataacagct gaaactcagc cagctatg                           1778
```

<210> 39
<211> 598
<212> DNA
<213> Homo sapien

<400> 39

```
caccccacag gagtaactca tcaggactta ccagactgct gcttttgggc atcatctgct     60
gggttgatga tttggtttgg ccaagagtct tgccaagact ttaatctatg cctcttgttc    120
tacatgaatt cttgggaatt actcacgttc cataggaaga gtgcatcccc aggtgatggt    180
ttttggttat ggtatgatcc tttcacaccg agggatttca ttgtttaaaa cgtgtttctt    240
taaaagaagc cttgataacg agagtggggg aaggaggcag cagactttga agactgtggc    300
```

ctttggtgtt ctggagtagg gggagggaag gagaaacatg ttttccacat catcgcaagt 360

gtgtgccctt tgccccttttt caggatcctt agagttgcct ccctccctcc accccgacag 420

ttttgcaata atgtgcctta tcagttgtga gtttacaggt gaagcaattt cccaaataaa 480

tggatgtaag tgttcaaaaa aaaaaaaaac aaaaaaaaag gctgggggaa accggggcca 540

aagcctctcc ccggggggac attgtttccc gccccaattc aacccacaca aaccaccg 598

<210> 40
<211> 2910
<212> DNA
<213> Homo sapien

<400> 40
gtccgcccgc cgctgcgtcc cggagtgcaa gtgagcttct cggctgcccc gcgggccggg 60

gtgcggagcc gacatgcgcc cgcttctcgg cctccttctg gtcttcgccg gctgcacctt 120

cgccttgtac ttgctgtcga cgcgactgcc ccgcgggcgg agactgggct ccaccgagga 180

ggctggaggc aggtcgctgt ggttcccctc cgacctggca gagctgcggg agctctctga 240

ggtccttcga gagtaccgga aggagcacca ggcctacgtg ttcctgctct tctgcggcgc 300

ctacctctac aaacagggct ttgccatccc cggctccagc ttcctgaatg ttttagctgg 360

tgccttgttt gggccatggc tggggcttct gctgtgctgt gtgttgacct cggtgggtgc 420

cacatgctgc tacctgctct ccagtatttt tggcaaacag ttggtggtgt cctactttcc 480

tgataaagtg gccctgctgc agagaaaggt ggaggagaac agaaacagct tgtttttttt 540

cttattgttt ttgagacttt tccccatgac accaaactgg ttcttgaacc tctcggcccc 600

aattctgaac attcccatcg tgcagttctt cttctcagtt cttatcggtt tgatcccata 660

taatttcatc tgtgtgcaga cagggtccat cctgtcaacc ctaacctctc tggatgctct 720

tttctcctgg gacactgtct ttaagctgtt ggccattgcc atggtggcat taattcctgg 780

aaccctcatt aaaaaattta gtcagaaaca tctgcaattg aatgaaacaa gtactgctaa 840

tcatatacac agtagaaaag acacatgatc tggattttct gtttgccaca tccctggact 900

cagttgctta tttgtgtaat ggatgtggtc ctctaaagcc cctcattgtt tttgattgcc 960

ttctataggt gatgtggaca ctgtgcatca atgtgcagtg tcttttcaga aaggacactc 1020

tgctcttgaa ggtgtattac atcaggtttt caaaccagcc ctggtgtagc agacactgca 1080

acagatgcct cctagaaaat gctgtttgtg gccgggcgcg gtggctcacg cctgtaatcc 1140

cagcactttg ggaggccgag gccggtgatt cacaaggtca ggagttcaag accagcctgg 1200

ccaagatggt gaaatcctgt ctctaataaa aatacaaaaa ttagccaggc gtggtggcag 1260

gcacctgtaa tcccagctac tcgggaggct gaggcaggag aattgcttga accaaggtgg 1320
cagaggttgc agtaagccaa gatcacacca ctgcactcca gcctgggtga tagagtgaga 1380
cactgtcttg acaaaaaaaa aaaaaaaaaa aagagaaaaa aagaaaatgc tgtttgtatt 1440
ttgtggtcta ataaggagtt cgggatagcc tgttgtattt gcctcatgcc agcccctgag 1500
ctgccttggg agaagatgct gattgtcctt gtccagagta ctgcttttgc agagtgacag 1560
gctgctggga cagatgtcct cctgttgcat ctttgtggat gtttagtacc aatgatgaca 1620
cgggaactca catcactgac accgctcttc atcttctgtt agtctcttga agagcatttt 1680
tttgtacttc tttgctgatg acctacctct tcataagcca agtgaaacaa gttgacgaac 1740
tgcctaggac ttccacgtgt tgctcacata catgatgatt tctgtcacgc tcttgtgttc 1800
agacacactg acattaccat gtatgtcaga cctccttatg atcgcatgtc ctgacagtta 1860
agctgattgc aaacagacta ttaaatatga atggagcaaa cgctgtatgt catggatatg 1920
ttctggagaa attcttaccc atctggatgg ggcagggccc ttgactcacc tgaatcatga 1980
ccaggcaaac attttatctg tcctttctgc aggaatccgt tctgtgtcat gctaggagaa 2040
tgggttcagt atatggggcc atcaggcagt ataccctctg aatgtttttc attgttgtat 2100
ttgcttagag taactaaaca attgtatctt ttaatttatc ttttaattca aagaggaaac 2160
cttggcttct gataactttg ttgtgttgta tcttaatggc ctatagctgt cattacttcc 2220
tgtagctgca gtacagaatt gttacagacc tggattaatg cttccaaaga cagaaggacc 2280
ttggcaccta aactgaccag ccctgtgatc ctgcacccca caggagtaac tcatcaggac 2340
ttaccagact gctgcttttg ggcatcatct gctgggttga tgatttggtt tggccaagag 2400
tcttgccaag actttaatct atgcctcttg ttctacatga attcttggga attactcacg 2460
ttccatagga agagtgcatc cccaggtgat cgtttttggt tatggtatga tcctttcaca 2520
ccgaggattt cattgtttaa aacgtgtttc tttaaaagaa gccttgataa cgagagtggg 2580
ggaaggaggc agcagacttt gaagactgtg gcctttggtg ttctggagta gggggaggga 2640
aggagaaaca tgttttccac atcatcgcaa gtgtgtgccc tttgcccctt ttcaggatcc 2700
ttagagttgc ctccctccct ccaccccgac agttttgcaa taatgtgcct tatcagttgt 2760
gagtttacag gtgaagcaat ttcccaaata aatggatgta agtgtaaaaa aaaaaaaaaa 2820
aaaaaaaaaa attctggggg aaaccggggc caaagcctct ccccgggggg acattgtttc 2880
ccgccccaat tcaacccaca caaaccaccg 2910

<210> 41

<211> 369
<212> DNA
<213> Homo sapien

<400> 41
cgtggtcgcg gcgaggtgag caactttttc aactatttta attgccaagc aagacatacg 60

atgccgtctt aaaatattcc cgcacgcaca gtccactcag catccagaaa ctgagcctcg 120

tttccgatct gatgatactt gtcattcaac tggtgggaga cccatttatt ttgtccccag 180

ggtctctaag gatgccgctg ccgctttctg atagctagat ctcaactctt acaaactttt 240

actgagcctc agattcccgc ttgcactact ccagctgttg ctgccgcagc tgcctttccc 300

agacagaagc gctgcgttgc aattcttcag ttacaatctg ggtaataaag gaaatcatat 360

ttaaaaccc 369

<210> 42
<211> 1236
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (27)..(27)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (1057)..(1057)
<223> a, c, g or t

<400> 42
cagtagtggc tagccctttt caacgangca taagtagtga agtgaaagct aaaggcttcg 60

cgccgcaact cactgcccga taacttctga agtgaaagca gaacacaaaa ccttcgaatt 120

atctggagca actttttcaa ctattttaat tgccaagcaa gacatacgat gccgtcttaa 180

aatattcccg cacgcacagt ccactcagca tccagaaact gagcctcgtt tccgatctga 240

tgatacttgt cattcaactg gtgggagacc catttatttt gtccccaggg tctctaagga 300

tgcgcgctgc cgctttctga tagctagatc tcaactctta caaactttta ctgagcctca 360

gattcccgct tgcactactc cagctgttgc tgccgcagct gcctttccca gacagaagcg 420

ctgcgttgca attcttcagt tacaatctgg gtaataaagg aaatcatatt aaaactacag 480

atttttgtat tttaaagcgc gatttccttc cccgagtag cgcgtggcgg agatggttgg 540

tgacccctgc ggtgcaatca tccaacgaaa acttcactgc tctgcaccag tttttctttt 600

tctgggttta gaaggggagt tgtgactaca atactttaaa aaaacaaaac tacctttatt 660

gtggcaagag aaaacattac aatggaaaca aattacatat aaatagttgt tttctcggta 720

gcttgttgtt ctttttgtct tttgtaaatt tctttggtgg tcttaagtcc tgatttggcc 780

ttgactacag tacaggagcc catggaccta gaccagagac cacggaccct gaaaataccc 840

actgaatcta caccgagatt gaatgcacaa cacagtttgc aacttgtaag gcttgtattt 900

tatacagtag ttgaaaccga taaattcgtt gtaaatcatt tgggaaggtt aaaaaattta 960

attaaaaata aaatgaaagg caggaagatt gacatttgta gtgcctaacg cccctgaatt 1020

ttatttccat ctttaagaaa agaaaaccaa aagccanacc gactctggca gcgcggggag 1080

acttcccagc cgttttcttg ggcagctagg acagcacctg ggccgtgcca gaaatgccag 1140

ttgctatggt cctgatggat ggaaaggaag tgcaggttgg gtctgcgcag catgtctaaa 1200

gcctgaaata ccgaagccac ccttacctta cttccc 1236

<210> 43
<211> 347
<212> DNA
<213> Homo sapien

<400> 43
gtactggata gcggccgccg ggcaggtcag ggagagcata gatatatccc tttttcagct 60

cagaggaaaa cacccaacaa caacctcaac aaacacaaca acaacagcaa aaggacgaag 120

cgtgagccag aaattccatc tgcgtgagta tgctgggaaa agcctggcga ggaatcctgg 180

ttggtgaaaa acagattcgg tgcttattgt tctgctctgt gagtaagtca ccaaagaaat 240

gtgggcgtgt gctgctggaa agaaaataga cttccacaaa aagtatgaat cactttgtta 300

attgtcaaag aaaatatata ttaagtgatt acttagcttc tggaatt 347

<210> 44
<211> 539
<212> DNA
<213> Homo sapien

<400> 44
cggacgcgtg ggcggcgggg caggacatct cccagacagc tggggacacc ctttccagcg 60

ggagcgggag gcgctgttgt tcaggaaagg aagttgtgtg tttgtgaatg agcgcacttt 120

gtggttcatg tgtttgggca ggacaaacta caccacaggg ttgggcgggg ctgtgtgtgc 180

tgggcacgta cgcttctttg ttttgtcagc ttcccgagag atcagggaga gcatagatat 240

atcccttttt cagttcagag gaaaacacct aacaacaacc tcaacaaaca caacaacaac 300

agcaaaagga cgaagcgtga gccagaaatt ccatctgcgt gagtatgctg ggaaaagcct 360

ggcgaggaat cctggttggt gaaaaacaga ttcggtgctt attgttctgc tctgtaagtc 420

accaaagaaa tgtgggcgtg tgctgctgga aagaaaatag acttccacaa aaagtatgaa 480

tcacttttgt taattgtcaa agaaaatata tattaagtga ttacttagct tctggaatt 539

<210> 45
<211> 449
<212> DNA
<213> Homo sapien

<400> 45
ccgcccgggc aggtacagtc cagcctcggc tgggcatcag agggaaacca tgcaaagagg 60

gggaggggga gagggagcca atattttgaa atttattgag acattttaga acctagtata 120

tagcctcttg gtgaatgttc tgtgtgttct taaaaagtga atgtgtattc taccactgtt 180

cagtaaatgc caattgggtt aagtttgttg atagtcaaat ctatatcctt acccatcttt 240

ttgtcccatt ttttctgtca gttattgaac aagaggtgtt aaaatctcca attacttcta 300

tttctctaac actgccattt ttttctttgt ggattttgaa tttctctata tattttgtat 360

attttgaagg tcacatacat cttttgtcgt catgtattct gatgaattga ccctttttgtc 420

attattaaat gtgctttatc tctattcat 449

<210> 46
<211> 598
<212> DNA
<213> Homo sapien

<400> 46
aggagctgga gacagcccgg ccaacacggg gaaaccccgt ctccgccaaa agatgcgaaa 60

accagtcagg cgtggcggcg cgcgcctgcg gtcccgggca ctcggcaggc tgaggcagga 120

gaatcaggca gggaggttgc agtgagtcga gatggcggca gtacagtcca gcctcggctg 180

ggcatcagag ggaaaccatg caaagagggg gagggggaga gggagccaat attttgaaat 240

ttattgagac attttagaac ctagtatata gcctcttggt gaatgttctg tgtgttctta 300

aaaagtgaat gtgtattcta ccactgttca gtaaatgcca attgggtcaa gtttgttgat 360

agtcaaatct atatccttac ccatcttttt gtcccatttt ttctgtcagt tattgaacaa 420

gaggtgttaa aatctccaat tacttctatt tctctaacac tgccattttt ttctttgtgg 480

attttgaatt tctctatata ttttgtatat tttgaaggtc acatacatct tttgtcgtca 540

tgtattctga tgaattgacc cttttgtcat tattaaatgt gctttatctc taaaaaaa 598

<210> 47
<211> 255

<212> DNA
<213> Homo sapien

<400> 47
accttattac cataaatcat cttgatgctg ctgataagat tctatttgct tttctttatt 60

catagagacc acaaacagat cgcagatcca ggtttctcaa actggagcat ctgcttaatt 120

ttcccataaa atcagtctta ttctttctga cagctctgag actcctccgg ccacgactag 180

gtgctgtcct ggaggaaacg gtggaggacg gccgcacaaa aaccaatcta cctgatgaaa 240

actccgttcc cttct 255

<210> 48
<211> 1403
<212> DNA
<213> Homo sapien

<400> 48
ggccgcattt tttttttttt tttttagaca gtctctctct ctgtcgccca ggctagagtg 60

caggcagagt ctgagcatcc agatttcaaa attaaaaaat aaaagataat ctagtttaat 120

atatagtagt tgaatcacct taagtctaga ctgctgtatg agcacccatt atctttcact 180

atattccatc atccccccaac atatccacag tagatgaagg gcagtttgct caaacattgt 240

tttgatcctg tcatgtctgt tcagaaatgc ctgtctattc agaaacccac gtctaataac 300

aaaatcttgg actggttact atcaaaaccc aacaacatac agactcctca gctaggccct 360

agggatattt ttctaccttg atttccaaat gttcattgaa agaatgctta attctaattt 420

ggaaaaaagt ttttggcttc ccacttctgc tttacacgtt catctttctt gaaatcaaat 480

ccaatccaat ctatattcta agaacctgct caaatcttgg ttcttcaaag ctttccctgg 540

tattttgcat ttttgctttg aatagttcca cgaaggataa cttcttactc cttccttcat 600

ctttctgtat cttgcatata gtaaatatta atgacttgtt tgcattttgt tatcctaact 660

tggctataaa gaaaatcaga tgtcttcacc agtcgttcaa acttcaggtc tgcctacaga 720

ttcatagatg gctgtggatt tttataattt tgtcacaaag ttagtggtaa ctacaggtta 780

tctcagaata tctttttttgg cgtataaatt tttttctttt cctttttttag acagtctctc 840

tctctgtcgc ccaggctaga gtgcagtggc gtgatcccgg ctcactacaa cctctgcctc 900

ctgggttcaa gagattctta gcctcagcct cccgagtagc tagggttaca ggcgcgcacc 960

acctccatgc ccagctcttt tgtattttta agtagagaca gggtttcacc atgttggtca 1020

ggctggtctc gaacttctga cttcaggcaa ccttagccgc ctcggcctcc caaagtgctg 1080

ggattacagg cacaagccac tgcacccagc cttattacca taaatcatct tgatgctggt 1140

acctgataag attctatttg cttttcttta ttcatagaga ccacaaacag atcgcagatc 1200

caggtttctc aaactggagc atctgcttaa ttttcccata aaatcagtct tattctttct 1260

gacagctctg agactcctcc ggccacgact aggtgctgtc ctggaggaaa cggtggagga 1320

cggccgcaca aaaaccaatc tacctgatga aaactccgtt cccttctcgc cagaaacata 1380

aaatgcgatg gatacgctcg tgc 1403

<210> 49
<211> 469
<212> DNA
<213> Homo sapien

<400> 49
ccgcccgggc aggtctcctg acctcaagta actcactcac ctcagcctcc caaagtgcca 60

agtgaaatat tttagtggga aaatgtccac agaataaggg gtgtgggctt ggtcagtgga 120

gagtggttga agaggtgaga gggaaccaag gctcttcctg ctatttcttt gctagggctg 180

gcccctttag gggcttccct tcccttagag atgactgcca gggtcttcca tccctacttt 240

ctgccccttc tctgaggggc ctgagccccc tcagcctccc aggttcatcc aaacaaacaa 300

cttatggcct tcataaaact gtctgattaa atgtgtatta ttatttttcc tttttattat 360

acaggatata aatgcttaac tcttttcatg tatttttaaa ccaactgaca aataattttg 420

agcttgtcat tagcatcctt gggctaatat aataaaactt tgttacttg 469

<210> 50
<211> 479
<212> DNA
<213> Homo sapien

<400> 50
ccgcccgggc aggtctcctg acctcaagta actcactcac ctcagcctcc caaagtgcca 60

agtgaaatat tttagtggga aaatgtccac agaataaggg gtgtgggctt ggtcagtgga 120

gagtggttga agaggtgaga gggaaccaag gctcttcctg ctatttcttt gctagggctg 180

gcccctttag gggcttccct tcccttagag atgactgcca gggtcttcca tccctacttt 240

ctgccccttc tctgaggggc ctgagccccc tcagcctccc aggttcatcc aaacaaacaa 300

cttatggcct tcataaaact gtctgattaa atgtgtatta ttatttttcc ttttttatta 360

tacaggatat aaatgcttaa ctcttttcat gtatttttaa accaactgac aaataatttt 420

gagcttgtca ttagcatcct tgggctaata taataaaact ttgttactta aaaaaaaa 479

<210> 51
<211> 312

<212> DNA
<213> Homo sapien

<400> 51
cgaggtaatg cttcggtgtg atgacagcgc acgttaacct cgaattcctg ggctcaggtg 60

aatcctccca cttccgcctc ctgacttgct gaaactacag gcacccgcct ccaccgccag 120

ggcccagccc acagctcctt tgacctcagt gacaggcact cacctacctg acccccaaac 180

tgaagcctca cttttcccag ccgtgcccac accctctggg ctaccccatt accatgacaa 240

gtattccctc tgctccagga gaaaagccag gtcccagacc tgacccatta aaacccaatc 300

attccagctt tc 312

<210> 52
<211> 568
<212> DNA
<213> Homo sapien

<400> 52
agcgggtggg ttctgggcca gcccagcctg gaggaggtgt gagaggctga gccactgctc 60

agcttagcgg ggggaccact tagtgaccaa caccctgagg gaggccccag catcccctac 120

ttagcttggc agcagcagcg ggataaatag gggggcactg ctgcctgtga gccagcccag 180

catagccatg ggtgtgtggg ggaagcagac agagacaggg tcttgctccg ctgtccaggc 240

tggaatgctt cggtgtgatg acagcgcacg ttaacctcga attcctgggc tcaggtgatc 300

ctcccacttc cgcctcctga cttgctgaaa ctacaggcac ccgcctccac cgccaggccc 360

agcccacagc tcctttgacc tcagtgacag gcactcacct acctgacccc caaactgaag 420

cctcactttt cccagccgtg tccacacccct ctgggctacc ccattaccat gacaagtatt 480

ccctctgctc caggagaaaa gccaggtccc agacctgacc cattaaaacc caatcattcc 540

aaaaaaaaaa aaaaaactct ccagcgct 568

<210> 53
<211> 294
<212> DNA
<213> Homo sapien

<400> 53
gaaaataagt cgctgtttgg ttaagtaatt taggagcaaa gcaatgctcc aagcgaggcc 60

tcctgcttca ggaaagaacc aaaacactac cctgaagggc cagcctagcc tgcagccctc 120

accttgcagg gagccttccc ttgcactgtg ctgctctcac agatcggtgt ctgggctcag 180

ccaggtggaa ggaacctgcc taaccaggca cctgtgttaa gagcgtgatg gttaggaaat 240

cccccaagtc atgtcaactc tcattaaagg tgcttccata tttgagcagg cgtc 294

<210> 54
<211> 779
<212> DNA
<213> Homo sapien

<400> 54
agtggctctg cggggaattg tgaggcggac tgcgagggaa agggggcctt gttgagtccg 60

ccaggattct accatcagaa aagaggccaa acttctatca tcatggtgga tgtgaagtgt 120

ctgagtgact gtaaattgca gaaccaactt gagaagcttg gattttcacc tggcccaata 180

ctaccttcca ccagaaagtt gtatgaaaaa aagttagtac agttgttggt ctcacctccc 240

tgtgcaccac ctgtgatgaa tggacccaga gagctggatg gagcgcagga cagtgatgac 300

agcgaaggtg ggctgcaaga gcaccaagca ccagaatcac atatgggact atcaccaaag 360

agagagacta ctgcgcggaa gaccagacta tcgagagctg gagagaagaa ggtttcccag 420

tgggcttgaa gcttgctgtg cttggtattt tcatcattgt ggtgtttgtc tacctgactg 480

tggaaaataa gtcgctgttt ggttaagtaa tttaggagca aagcaatgct ccaagcgagg 540

cctcctgctt caggaaagaa ccaaaacact accctgaagg gccagcctag cctgcagccc 600

tcccttgcag ggagccttcc cttgcactgt gctgctctca cagatcggtg tctgggctca 660

gccaggtgga aggaacctgc ctaaccaggc acctgtgtta agagcatgat ggttaggaaa 720

tcccccaagt catgtcaact ctcattaaag gtgcttccat atttgagcag gcgtcaaac 779

<210> 55
<211> 184
<212> DNA
<213> Homo sapien

<400> 55
acggcagacg aaggcagtcg aggtgtggag gtgatcacga agatacatgt gtttgactgt 60

ttaatttgaa agtttacatt ttttatgctt tgtgttggtg tgtaattttg gtactcttgg 120

tggctagttt ttgtcaaatc ttttttggaa tattgcttaa agtgtttttg attttatgat 180

agtg 184

<210> 56
<211> 2065
<212> DNA
<213> Homo sapien

<400> 56
tttttttttt taaacagatg gagttactgt gaagaagttt tcacaactat ttatgctggt 60

aaaacaaatg ctgttaaatc accttatgcg tcgttaacta cttcattggg gctaattacc 120

cggaatacgg tctcaccgat gcagttttca tggacataga aaattcaaat agaatatata 180

atattgaatt taagatttgg ggggttaaaa aagaaaactt aactttataa aattatttat 240

tctattttaa gccttctatc atattttccc atccaattgt ttggtttcag tggtccagct 300

ttatttacag gcatataaaa tgaaattgtg agatgttttg caagcttctt tttactttga 360

gtagctttta atttgtatgt ttttatgtgg atgaagagca ttttttatgc ttttgtgcaa 420

taggttccaa tatgcattta ttagacatct gtttaaatgg taatgtagca tttattttgc 480

taaattgaaa gggaacatag atggaattcc aaaatatgta cattcagctg tttggttttt 540

cgtttttcat tgttattatt gtgagaatgc tgttattggg gttgtgtgtg agtgcccgtc 600

agccagtgat gcctcgggcc acgctgtggg gccacctcag tcctgcctgg gtcctggtgc 660

cttggacccc acgtgcttgt ggccaggctg cccctgggcg gggccatgtg gcctcagacc 720

acaagagcgg agctgccctg gcccaagcac tgcagctgcc tgcacccccg ggcttcgcag 780

ccttgcttgt tttctctgaa cagcaacaga acagtgttca cagcgattca aagggtggca 840

ttgggttgga cgttctgggt acaagccaac ctagtcccac gttgtacgtg aatgtttaat 900

gtgctctcaa aacatggaaa ataagtttag tgcacatagc taaatcacaa aacatccaat 960

ttctctgttt cctcaggaag tcattactgc gccaccacat cacatgacct taacatgatc 1020

aatgtatttc tctgccttga catttaaata cataaattga gataagtaga ttagaaaatc 1080

attcaaatga taccataatt tgtacgggac agggtgcggg caatggccac gtggccaagg 1140

ccccgcagga acgcgccgag gtctccctca ccctccaggt gtccttcgca cccaacagtg 1200

cgtctgagga acgagctgca gtttgagcgt tcccctgaga tgtgcgtagc ctccgtgtaa 1260

atgtccactc ccatggctta attgcctatc agacgcattt tcccagacga aagcaatgtt 1320

gggttgggga agacagtgca gccacccagc ctttaccagc agcgtacggc agacgaaggc 1380

agtcgaggtg tggaggtgat cacgaagata catgtgtttg actgtttaat ttgaaagttt 1440

acatttttta tgctttgtgt tggtgtgtaa tttttgtact cttggtggct agtttttgtc 1500

aaatcttttt tggaatattg cttaaatgtt ttgattttat gatagtgaag cttgtattca 1560

gtgttttgcc aattaatatt atatgcttgt aataaaagca aaagaaaagc ttaagtgaaa 1620

gtgtttctgg ctggaaatgt attacataaa tcatttgtat gagcatgaga gatgacttgc 1680

tcatgggatg agcaatgctg tgggtattgg atggaccact caggcagcct ggccgcacac 1740

acacacatgc atgctacata tgtgtacgct cagaccacat gcctgcacat atgtccacac 1800

accagcccac ggctcccttc ctcaccccag ccctccccac tcccaggtta accaatgtct 1860

cgatctgctt caggcggcat ttgtaaactg tcaccctgct tgcaaaccct catttaaatt 1920

ggtccagcat ttgctacctg agtttaggcc acaaactttg taaagctcag tttgtctccc 1980

ttgcagaaag aggttgctgc agtttggagc tggttatgcc aaggaaaata cggctaaaaa 2040

agaacttact gattgccata ggtgt 2065

<210> 57
<211> 976
<212> DNA
<213> Homo sapien

<400> 57
agatgctgct cgaccgcgcc gataatgtga tggattttag taaacagcaa aatattttga 60

aagctggatg cagatgctca gatgctagag ggtgaaatgg acagacttgg ctaggaagag 120

atatgtgaat gttagcagag ggacctttct ggggattaag gaatcaggaa cacaaatttc 180

tcttctttcc ttcccaccag gctccatgcc ccccttactg gaggaccaag accttgttgc 240

cttcaattta cgggatccca gtgggatcct gatattttcc atagtttctt aacaacattt 300

caagttaaat attaaaatta ttcatagggt gtggagtgag ccaagtgcaa cacattgctg 360

tcaggggtgt tggctactcc gccagctgtt gaaaaaagga gaaagaaaga gagcaaactg 420

agatccacac accccacaca gtatgaccaa ggcgccttct gacttcagga aagccaggca 480

gacggggatc cctggatgct cacagcttgg cagccgatat tcactggagc cagaacagtc 540

tgctctgagg cttgtctgca tccagaagtt gcaggaaagt tccacaacgt gtgaagactt 600

cttttgtcct ctctgtggga gagctgggga acataggatt ccttatagac ttatcctccc 660

cacctctcca atgagcaaag gctgctaaaa acttctgaag cctgaatccc aaaagctgga 720

ggctttctct ctcctcccag tgatcggagc ttctcagggt ggcggattgt ctcatggttc 780

tgggggccca aggcaggttc ccggcaggag tgagggtccc ggctcttggg agaggccttt 840

cagctcttgg ctccggggtt ccacctcagc ggggctccct gcgctgcgtg ccggggtgcc 900

ggctccagac tggtcctctt cccccgcgct ttagctgggg actggccacg tccggttatt 960

tcccacccca aagaga 976

<210> 58
<211> 1660
<212> DNA
<213> Homo sapien

<400> 58
gccgtaccca cacctgcggt tgtgggctac tccaaacctg accggttaca tacttctcaa 60

taggcaggta ggtctcagtg aataattgaa aaatggtcct cactatgcac aggctataaa 120

gggactagta tttcagttta atcatgagaa aggctctctg ctttaagaag cactctacat 180

gcgtttccca agccaaactg ctggggctca actgtgggtt ctgctccttc ctggctgccc 240

aaactcaggc aagtgacttt atttcctcca tttcctcatc tataaaatgg gagtagttgg 300

gtggcattaa accgtctatg gaaagtcctc agcatggcgc ctggcacaaa gcaacggctc 360

cctcgcctct gttctcactc ctgttccaca gaattacacc cactcttctc tgctgatact 420

cttaatctca gatccccaat tctgtctttc aaatgtgttg tgtcaacatt tatttgcaaa 480

catgtctatt tgatttcaaa tgaaaacgct tttgagtgga tttaaaaaca aaacacatgc 540

gggggagaaa agagaggctg acagacatgt tagtaaacag cgaaatattt ttgaaagctg 600

gatgcagata gctcagatgc tagagggttg aaatggacag actttggcta ggaagagata 660

tgtgaatgtt agcagaggga cctttcttgg ggattaagga atcaggaaca caaatttctc 720

ttctttcctt cccaccaggg ctccatgccc cccttactgg aggaccaaga ccttgttgcc 780

ttcaatttac gggatcccag tgggatcctg atattttcca tagtttctta acaacatttc 840

aagttaaata ttaaaattat tcatagggtg tggagtgagc caagtgcaac acattgctgt 900

caggggtgtt ggctactccg ccagctgttg aaaaaaggag aaagaaagag agcaaactga 960

gatccacaca ccccacacag tatgaccaag gcgccttctg acttcaggaa agccaggcag 1020

acggggatcc ctggatgctc acagcttggc agccgatatt cactggagcc agaacagtct 1080

gctctgaggc ttgtctgcat ccagaagttg caggaaagtt ccacaacgtg tgaagacttc 1140

ttttgtcctc tctgtgggag agctggggaa ataggattcc ttatagactt atcctcccca 1200

cctctccaat gagcaaaggc tgctaaaaac ttctgaagcc tgaatcccaa agctggaggc 1260

tttctctctc ctcccagtga tcggagcttc tcagggtggg gattgtctca tggttctggg 1320

gccaaaggca gttccaggaa ggaggtgagg gtccgactct ggagagaggc atttcagctc 1380

tttggctcag ggttccatc cttcagcggg gcatccttgc agtctgctgc ctgggtgccg 1440

gtctccagac ctggctctct tccctcgcgc tcttcttcaa gcttctggga ctcagctgcc 1500

accgtggtgc tcttgctgga gtggcacccc atctcagagg gacacagagg atccagtgcc 1560

cttggatgtt ccaggaggag gaaggtctgt gccttcctcc ttgggggcca gcgttaaata 1620

accatcctct tgcagcactg ttgaaaagag ccagttccgt 1660

<210> 59
<211> 686
<212> DNA
<213> Homo sapien

<400> 59
cgcccgggca ggtcaaccct agctgtgctt ctgcgagaag aaagggtgta gcaacttctg 60

gcagatatga ggcttttttc ttttttttt tttttttgga aacaacgttt ggctctgtca 120

cccaggctgg agcgcagtgt ggccaatctc gtgcccactg cagcctcgac ctctggtggc 180

tcaaggagat cctcaatgcc tcaagcctca ttatgtagct tgggcgccac agaacttgca 240

ccaccacacc tggctcgagc gccattttag tctctgaggt tgagcagctc aggagccggc 300

ctccagcacg gtgctgtgtt tgtgaaccag agaagggggc ccccgcgagg accccagaca 360

gggccttagg actctcatat cttctgtgtc ttctccatct gggggctctg gctctcgggt 420

ttcctctacc ttttcatggc cccagaattc catattgcgt aaaggaactc ttctgggata 480

taacatatta aagtgtcctc cctcagcaca aaaaaaaaca aaaaaaaaaa aaacgctctg 540

ggcggggaag ccctgtggcg ccaaaagcgc tgtgtcgccc ctgtgtgtgt agacattttt 600

tttctcccgc gccccaaatt cccccccaat aattcttcaa caacaaaaaa ttaaaacaca 660

aaaacaaaaa aggaaaaaac aaaacg 686

<210> 60
<211> 624
<212> DNA
<213> Homo sapien

<400> 60
ttgacatatg gaactgcgtc tctatgctgt cggcggcgca gtgtgatgga tgtcgcgcgg 60

gtcgatttac ctgtagtgtc ttatcactct ttcatgtcac aatagcgtgg agcattagag 120

aaacagccta gactttcagc ttgatagcca gtgtgaaata tcattgatag aattttagtt 180

gttaggaaaa attggtttga tttctagctt tattcactat taggtatgtg agcttgggca 240

aatcgcttat atctttgagt ctagttttct cctcagaatg cagaacatta ggctaaatga 300

tttccgagtt tccagcgtag tcctagagtt ctatattttc tacgatagtt gaattattta 360

tatcatgctg ttgctgggga atatgactaa ccccttttgaa gctactaatt ttatgtcgag 420

ctttaagagt ccaatagttg ttatcttcag aaaatattat ttgacctaca gtatgtccaa 480

catcaattta ataaaatcgc tttataacag caagaaaact taaccaaaat taaaagacta 540

agaaagggtt gggggtcact cggggccacc gggtgtcccc tgggtgacct gttttcccgg 600

cccaatttcc caacttagag caaa 624

<210> 61
<211> 913
<212> DNA
<213> Homo sapien

<400> 61
ctggaagggg ttttttgtat caactaacaa cctcagcgca taaaggagat ttaaaaggag 60

cacatgattt agtgggtggg ccatgaaact agagatggga tttgggggtg aatttgtcca 120

atatcatgga ttttaatcca gacatctctg ctaacaagcc tttggtaagt cacttcagat 180

acttttcctc ctttttacaa agagagggct ggcttagtta tttgccaaag ccccttccag 240

gcctgaattc cacaagtacg atttactgta gtgtcttatc actctttcat gtcacaatag 300

cgtggagcat tagagaaaag cctagacttt tagttgatag ccagttgaaa tatcattgat 360

agaattttag ttttaggaaa aattggtttg atttctagct ttattactat taggtatgtg 420

agcttgggca aatcgcttaa tctttgagtc tagttttctc tcaaaatgag aacattaggc 480

taaatgattt ccgagtttcc agctagtcct agagttctat atttctacat agttgaatta 540

ttttatcatg ctgttgctgg ggaatatgac taaccctttt gaagctacta attttatgtc 600

gagctttaaa gtccataatt gttatcttca gaaaatatta tttgacctac agtatgtcca 660

aatcaattta ataaaatcgc tttataacag cgctacaacc acacaaagca acaaaaacag 720

acccacaaac aacacacgtt tgggccggcc tctggcccca agaaaaaagt ttttaacaaa 780

ccacttccgt ggggacgcgg cgggcccaca gaaggaacgg gaacaacggg cccaaacggt 840

ggtccccaag aaactcgggg tacgagaaca ggcgtgggga ccccccaaaa aaaaagcccg 900

ggggaacccg gag 913

<210> 62
<211> 356
<212> DNA
<213> Homo sapien

<400> 62
cggatgatag actaatatag gcgaattggt catttagatc atgtcgatcg gcgtcattgt 60

atggactcgt ggtcgcgtgc cgattgtacc cccttccgag tacgacggct cttgtggaac 120

tgcgcggtcg atagcagcgt gctctaggag gagagtcaac gtcaggctcc aggggtttga 180

accgatccac tttcagttac gatgtatctg agaccagaga gcagaggggg cttatttccc 240

tctgccctat actggccgtc gaatgagcaa cagtcacaca gagcaggcga ccttttttgtc 300

aaaagtgtgt ggggcggggc gcgcagtagg cgccctaaac gctggactga acagag 356

<210> 63
<211> 829
<212> DNA
<213> Homo sapien

<400> 63
ggatgatacg acccacctat aggcgctgtt cctctagatc atgctcgagc ggcgccatgt 60

gatggatcga gcggcgcccg ggcgggtggg gcccactttt tatacaagcg ggtggcggtg 120

gatccaggcg cacagacggg agtgaggcgt ggcggagacg cgtgcacgct gggtctgtgc 180

acttttgctg ctgctgcttt tttttgggga gggccccttt tggccccccgg gtgggtgtcg 240

gggggactcg ggcctcatat acaccccttt gtcccccggg gtttagcgca gacatccccg 300

tgtcttaagc ctctcccact ttgctcttaa acgagacggg ccgcctgtga caccaattac 360

cccgcgctag ttttgggtat ttgaggacaa gatgggttcc ccccttgtgt tggcgttggc 420

ctggaccctc cttgccccta tgtgaatcca ccttaatgtg gcgcctcccc aaggttgcgt 480

gggattccgg gcgtgagtca acaagccccc ggtgggtttg ttggcctacc tataaccccc 540

ttttataact ttttggtaac catggggata caaagatgga atttcgcggg tcgcgaaaag 600

aattcgcagg gagaaaccta atattccctt tacagtgtgt tgttctacaa tgtgggtcac 660

ggtttttggc acaatcaatt caccattttt tggccaacgc ccaaaaaaaa aaaaaaaaaa 720

acgcctgggg gttaccgctg gcgcaaaagc attccccggg gtgagatgtt tcccgccccc 780

atcacaagca aagacaaaaa aaaaagcaaa aagacaaaaa aaaaaacat 829

<210> 64
<211> 982
<212> DNA
<213> Homo sapien

<400> 64
ctccaccgct cctcccaggc tcacgaggct cacgggctac ccagcgccac ggcccaggag 60

gctggaccac aggcctcagt cccgacttct ccctgctgcc cggcacggga ccctccccca 120

ggacagccac ctcccagtgc tgaagagccg gcccgcctca gagccacccg ccgttccttt 180

ttcctatgca tatacttctt tgaggatctg gcctaaagag actacgaagc tgcaccgtgt 240

ggggcccact ttttatacaa gcgggtggcg gtggatccag gcgcacagac gggagtgagg 300

cgtggcggag acgcgtgcac gctgggtctg tgcacttttg ctgctgctgc tttttttggg 360

ggagggcccc ttttggcccc cgggtgggtg tcgggggggac tcgggcctca tatacacccc 420

tttgtccccc ggggtttagc gcagacatcc ccgtgtctta agcctctccc actttgctct 480

taaacgagac gggccgcctg tgacaccaat taccccgcgc tagttttggg tatttgagga 540

caagatgggt tccccccttg tgttggcgtt ggcctggacc ctccttgccc ctatgtgaat 600

ccaccttaat gtggcgcctc cccaaggttg cgtgggattc cgggcgtgag tcaacaagcc 660

cccggtgggt ttgttggcct acctataacc cccttttata actttttggt aaccatgggg 720

atacaaagat ggaatttcgc gggtcgcgaa aagaattcgc agggagaaac ctaatattcc 780

ctttacagtg tgttgttcta caatgtgggt cacggttttt ggcacaatca attcaccatt 840

ttttggccaa cgcccaaaaa aaaaaaaaaa aaaacgcctg ggggttaccg ctggcgcaaa 900

agcattcccc ggggtgagat gtttcccgcc cccatcacaa gcaaagacaa aaaaaaaagc 960

aaaaagacaa aaaaaaaaac at 982

<210> 65
<211> 342
<212> DNA
<213> Homo sapien

<400> 65
tggtcgcggc cgaggtgttc ccatttattt cctgtggaac tgaatccctc ctccctccac 60

tccttgggag cccaggtggt ccttgaccac cattcaggct ttccaagaag ccaaccacct 120

tggagatttt ttttcttgaa tttcgctgtt ttcttctgct tcctttagat aaaaagcagc 180

tcaagagacc ttatcttagg gatgagaaaa acatgcatat taattccatc tgagtgattg 240

tcagtgtaag gccttttaaa acaaaagcaa gttctttgtt aggaattggt caaaattcat 300

ctctttcttt aagcccatca actcccagga cggtttgagt ta 342

<210> 66
<211> 5814
<212> DNA
<213> Homo sapien

<400> 66
tttgggggtg ataaaaaggg gggcccaaaa aacgggggag cggagatttt tttgggaaat 60

tttttttttt ttcctttgga tatatgacca gcagtgggat tgctggatct tacgatggaa 120

ttcccaaaga tgttgaccag gaagatcaag ctgtgggaca tcaacgccca catcacctgc 180

cgcctgtgca gcgggtacct catcgacgcc accacggtga ccgagtgtct gcacaccttc 240

tgcaggagct gcctggtgaa gtacctggag gagaacaaca cctgccccac ctgcaggatt 300

gtgatccacc agagccaccc cctgcagtac atcggtcatg acagaaccat gcaagatatt 360

gtttacaaat tggtaccagg cctccaagaa gcggaaatga gaaagcagag ggagttctat 420

cacaaattgg gcatggaggt gccgggagac atcaaggggg agacctgctc tgcaaaacag 480

cacttagatt cccatcggaa tggtgaaacc aaagcagacg acagttcaaa caaagaggcc 540

gcggaggaga agccggagga ggacaacgac taccaccgca gcgacgagca ggtgagcatc 600

tgcttggagt gtaacagcag caaactgcgc gggctgaagc ggaagtggat ccgctgctca 660

gcccaggcga ccgtcttgca tctgaagaag ttcatcgcca aaaaactcaa cctttcatcc 720
tttaacgagc tggacatttt atgcaacgag gagatcctgg gcaaggacca cacactcaag 780
ttcgtggttg tcactaggtg gagattcaag aaggcgccgc tcctgctgca ctacagaccc 840
aagatggact tgctgtgaat ggtgccacac agcgcccaca gactgggccc tcgcacccctt 900
gggtgctccc ggccgccgcg cttaagaaca ttgcctctgg gtgtcatgtg gaccagactt 960
ctgaatagag aatatttata acttttgtat gagagagaat tcacactcaa caagacacta 1020
ccagcaccac gtttacagag gatgaaaaca cttcacagtc tcccagagcc gatcgtcctc 1080
tcccccgccc caccccgtgc ttcagccttg cagggagagt gatgctccag gcaacacggt 1140
tctgagtcac cttctgacac gagctccctc tgcttgcttt ccaggtcttg aaaatctgaa 1200
ttcacttcag tttagtttat gaattttagg tttcatgata agcctcaagt gtagttggac 1260
ttttattgaa tccttcctaa gttattgaaa aaatgtcttt tcatggtgaa tgacaatatt 1320
tatgttgcct ttagcttctt gaagatttag aagttatata aaaaattaat ttaaaagcaa 1380
accaaaagag gtttccatta acattatgat ttaaccattg tatttaattt cccaccttat 1440
gaaacacaac agcagctccc tgactggttc gcctttcatt gtgtgaggtc ggcacttgga 1500
ctcactcaga actgtcgctc acctgtggct gacacaccca gccctggaaa cggggcccca 1560
gacgccacgt cgggatttct gacatgctca gcaggtagac cagaggccgt gtgaccagct 1620
cagtgctggt ttacggaaca actcttactt ttaaaaatta cttgttcccc caaattgttg 1680
agtgccgccg tttggtttcc tatgttttct ttccctgttt tgattttgct gaagggagag 1740
gtggtggtgg ttaggatcag agctctcctg gcatccgtgg ggaggatttg ctggtggtgg 1800
cttcgggctc atgcccagac acactcactg ccccgtctgt ccaaggcctc cccttcccct 1860
ttgctggtgg gaggagctcg tgtgctcctt ggccgcttac tggaagggcg tttttcagag 1920
ctgcagggac agggtgagca gctgaagggc taggagggaa gccggccccc gctctgcaga 1980
agctgcattt cagctgaatc tgtgtttcag cctcagttgg ttgcaccgtt agcccctctc 2040
ctcccggatg gtcatgtttt tgtcacatta gagaataaac agccacacac acattttttt 2100
ttttccttta aaacagtaac ttggaaatat gaaaaggcca gaaggaggag caagggctgt 2160
tttctggagt ggttgaggtg ttgtcctgca gttgtcattg tcttctccac cgggctgttc 2220
ccatttattt cctgtggaac tgaatccctc ctccctccac tccttgggag cccaggtggt 2280
ccttggccac cattcaggct ttccaagaag ccaaccacct tggagatttt ttttcttgaa 2340
tttcgctgtt ttcttctgct tcctttagat aaaaagcagc tcaagagacc ttatcttagg 2400
gatgagaaaa acatgcatat taattccatc tgagtgattg tcagtgtaag gccttttaaa 2460

acaaaagcaa gttctttgtt aggaattggt caaaattcat ctctttcttt aagcccatca 2520

actcccagga cggtttgagt tactcagtta cctaagcttg ctattcatcc aaatcatttt 2580

ctagagtcac tgtataaggg tctatgagta gctgtgtatg aataaatatt acctgtctac 2640

ctcaaaatac acatactgct gaagcattct gtacaaccgt gtgttatcac agtgcagttt 2700

taagtgtaac gttagaactt aggcattttc ctgtgtggcg gaataagaaa ggattaaaca 2760

gttacaagcc tccaaattca aataaaatta aatcacagtt cagatgaaac tgaatatcat 2820

tgtaataatc tcataatata tatttgtaac tttgtagcta tctttgaaat cacttgactt 2880

tgcaatggtg ctaagctgat agatttaaat acacagacgg gcgagtggcg cccgtgtcga 2940

tgtcttcagc cagtggtgac cctgcttttg taaccgcgtt aacctgacaa aacctcagca 3000

gcagaagtcc ctatttttct aggagttcat cgtgcagaca gtcttcacta caggactcgg 3060

ccctggggcc tctgcctctc gtctgacctt gcagccttag tcgttggagg ctggagcgca 3120

atggccctgc cgtctgtgga gcctctgggc ggccttcttt cctttctgtc aacctctcat 3180

ttcacagaaa aaggctgaat ttcatttttt ccagcatgaa agccaggatc ggttagtggt 3240

tggattctat tggttttttt tttaaacaga tggagttact gtgaagaagt tttcacaact 3300

atttatgctg gtaaaacaaa tgctgttaaa tcaccttatg cgtcgttttc aacagcagtg 3360

gggctaatta cccggaatac ggtctcaccg atgcagtttt catggacata gaaaattcaa 3420

atagaatata taatattgaa tttaagattt gggggttaa aaaagaaaac ttaactttat 3480

aaaattattt attctatttt aagccttcta tcatattttc ccatccaatt gtttggtttc 3540

agtggtccag ctttatttac aggcatataa aatgaaattg tgagatgttt tgcaagcttc 3600

tttttacttt gagtagcttt taatttgtat gtttttatgt ggatgaagag cattttttat 3660

gcttttgtgc aataggttcc aatatgcatt tattagacat ctgtttaaat ggtaatgtag 3720

catttatttt gctaaattga aagggaacat agatggaatt ccaaaatatg tacattcagc 3780

tgtttggttt ttcgtttttc attgttatta ttgtgagaat gctgttattg ggttgtgtg 3840

tgagtgcccg tcagccagtg atgcctcggg ccacgctgtg ggccacctc agtcctgcct 3900

gggtcctggt gccttggacc ccacgtgctt gtggccaggc tgcccctggg cggggccatg 3960

tggcctcaga ccacaagagc ggagctgccc tggcccaagc actgcagctg cctgcacccc 4020

cgggcttcgc agccttgctt gttttctctg aacagcaaca gaacagtgtt cacagcgatt 4080

caaagggtgg cattgggttg ggacgttctg ggtacaagcc aacctagtcc cacgttgtac 4140

gtgaatgttt aatgtgctct caaaacatgg aaaataagtt tagtggcaca tagctaaatc 4200

acaaaacatc caatttctct gtttcctcag gaagtcatta ctgcgccacc acatcacatg 4260

accttaacat gatcaatgta tttctctgcc ttgacattta aatacataaa ttgagataag 4320

tagattagaa aatcattcaa atgataccat aatttgtacg ggacagggtg cgggcaatgg 4380

ccacgtggcc aaggccccgc aggaacgcgc cgaggtctcc ctcaccctcc aggtgtcttc 4440

cccaacccaa cattgctttc gtctgggaaa atgcgtctga taggcaatta agccatggga 4500

gtggacattt acacggaggc tacgcacatc tcaggggaac gctcaaactg cagctcgttc 4560

ctcagacgca ctgttgggtg cgaaggacac ctggagggtg agggagacct cggcgcgttc 4620

ctgcagggcc ttggccacgt ggccattgcc cgcacccttg aatctccacc tagtgacaac 4680

cacgaacttg agtgtgtggt ccttgcccag gatctcctcg ttgcataaaa tgtccagctc 4740

gttaaaggat gaaaggttga gtttttggc gatgaacttc ttcagatgca agacggtcgc 4800

ctgggctgag cagcggatcc acttccgctt cagcccgcgc agtttgctgc tgttacactc 4860

caggcagatg ctcacctgct cgtcgctgcg gtggtagtcg ttgtcctcct ccggcttctc 4920

ctccgcggcc tctttgtttg aactgtcgtc tgctttggtt tcaccattcc gatgggaatc 4980

taagtgctgt tttgcagagc aggtctcccc cttgatgtct cccggcacct ccatgcccaa 5040

tttgtgatag aactccctct gctttctcat ttccgcttct tggaggcctg gtaccaattt 5100

gtaaacaata tcttgcatgg ttctgtcatg accgatgtac tgcagggggt ggctctggtg 5160

gatcacaatc ctgcaggtgg ggcaggtgtt gttctcctcc aggtacttca ccaggcagct 5220

cctgcagaag gtgtgcagac actcggtcac cgtggtggcg tcgatgaggt acccgctgca 5280

caggcggcag gtgatgtggg cgttgatgtc ccacagcttg atcttcctgg tcaacatctt 5340

tggcttctgg tcacagagac ttccaaagtc aaatcctccc cttctcctcc tgccctgcgg 5400

ggctgtgagg ttcgggagac ctgatgcata ctgaggacac gggagcctcc gctagacgtc 5460

cgcgtgtccg tccctctgcc ggggcatgga ggcttcaggc cagcactccc acgcgacgct 5520

gcctcctggc tccgcacgcg gactcgcttc tgcaccctcc ttccggccct ggctcttttt 5580

ccattttctg agcacgacgc ctggggtccc tgctacattg tgacgcctgc ggtctcgggg 5640

ctgcgggcga gcaaggcgca cgcggggtcc gcgccggagg ggggaggggg cgccgctccc 5700

tctgggtccc ccccaagagg aggagaaaaa agcccgtccg gagtttaaaa caaagaggcg 5760

gaaatgcccg agccgagccg agcgccgcgt cgggggcgcc gcgctcggag catc 5814

<210> 67
<211> 708
<212> DNA
<213> Homo sapien

<400> 67
cgtggtcgcg gcgaggtcaa cttggaactc tggaaatgtg gcttcgctca ctggcgcctt 60

gagcttgggc gactgccggg tccgcgaaac ccgacccctg cagagctgac tccgggacta 120

ttttagtttc taacgtcaac ttgccccgat tcaagagggt ttgcgcaaaa aacgtagccc 180

gttgtcctcc tgctgcagct gttgttgcag ctgtgtggct gcgtttagta ggaataacca 240

actcaaattg ggaagttctt cagctcagta tccgctcctg taattagaac ttcttttctt 300

taagcgatga aattttggac agagagatct ggagtttagt ttgtgacgtc gaagaacaaa 360

ctccaaaatg taataccttg tccccatttg gggggcaaag ttgtgggcta attcaattcg 420

ccatggaagt gtcttctttt taaagtagtt tagtaggtat atgaatgtat ctgtcagttc 480

ttgagagacc tatggattta gcagagattt taacttagtg ccaaaaagtt tcatatttaa 540

aggcgaataa agcgaatatt tcttaaaaaa aaaaaaaaaa aagggaaaaa aacaaaaaaa 600

aaaaaaaaag ggtgggggc cccggggcca aagggttccc gggggaattg ttctcccccc 660

ccatcacacc cacaacacaa aaaaatgaaa aaggcacaac cggaccat 708

<210> 68
<211> 1099
<212> DNA
<213> Homo sapien

<400> 68
tttcattata tattgctcta tattctaggt cgcccacttt acacttcctt ctcatgcact 60

tggtcaatac cacgccccgct gaccacactg gcgacttccc tctctgtcgc ccctccgtga 120

agtcagaccc actctgcggg ccaagaaagg tgaccgggct tccttccggc ttgctaagca 180

gaggccggaa gcggtggttt ttagcggatt ctcgtagctg tgccgggtga gtggcgctcg 240

cgttcgggcg cgtgagaccc atcccgggga cccgtctccg cgggggcagc tggagggcgg 300

cggggctcct ggcgccggta gcgcacctgg gcaggtgtgc cagccaggtc cccggttctg 360

ggatccgagg ccatggcttg gagtggccca gacccgaact tcgctcctgt gcccaaactt 420

ggaactctgg aaatgtggct tcgctcactg gcgccttgag cttgggcgac tgccgggtcc 480

gcgaaacccg acccctgcag agctgactcc gggactattt tagtttctaa cgtcaacttg 540

cccgattcaa gagggtttgc gcaaaaaacg tagcccgttg tcctcctgct gcagctgttg 600

ttgcagctgt gtggctgcgt ttagtaggaa taaccaactc aaattgggaa gttcttcagc 660

tcagtatccg ctcctgtaat tagaactttt tttctttaag cgatgaaatt ttggacagag 720

agatctggag tttagtttgt gactcgaaga aaaactccaa aatgtaatac cttgtcccat 780

ttgggggaa agtttgggct aattcaattc gccatggaag tgtcttcttt ttaaagtagt 840

ttagtaggta tatgaatgta tctgtcagtt cttgagagac ctatggattt agcagagatt 900

ttaacttaag tgcaaaaagt ttcatattta aaggcgaata aagcgaatat ttcttaaaaa 960

aaaaaaaaaa aaagggaaaa aaacaaaaaa aaaaaaaaaa gggtgggggg ccccggggcc 1020

aaagggttcc cgggggaatt gttctccccc cccatcacac ccacaacaca aaaaaatgaa 1080

aaaggcacaa ccggaccat 1099

<210> 69
<211> 770
<212> DNA
<213> Homo sapien

<400> 69
gaatatctca tatagcgcaa tggtcctaat catgccgagc ggcgcagtgt gatggatgcg 60

ccgggcaggt acctcttggc ctttgctggg cttgcgttct ctcctctagt ggtgtgggga 120

tgactttcaa tgactttcaa tacttcccct gaagcaagaa tgataacgga gaaatgtctg 180

tattgaggaa agggcttcga attccccagg tactgaacaa ttggtgtcgt gactgatgga 240

gaattctcag gagatgaatg agaaacgcct ttgcgaaagc tatgcaacag tttacatcac 300

gtcatgtgaa gctattcgtc taaagacacg agcaaatctg aagaccaaat tattctcctg 360

ttgaggtccg tggatggcag atttaaaggg aagaaccaca aaggcttgca aagataagga 420

gaggctccat ctctaatgca tgtagaagct ccttacgggt gcccatcacg agcatagctt 480

gtgaagccac cagtgctgtg cggaatctgc gtcatgcgcc gaagatgtca catgcaggat 540

ttccccaagc cagctccatc atcacagaca cggagagctg cagggcaggc ctgcccactg 600

ttttgtcgac tctggcctcc tctggcagca tagatcctta ggtgctcaat aaaggtgtgc 660

tgtattgaac caaaaaaaaa aaaaaaacaa aaaaaaaagc gctgggtac cggggcaaag 720

tgtcccgggt ggaaattgtt tcccgcccca aaatcccaat ccaaacacac 770

<210> 70
<211> 357
<212> DNA
<213> Homo sapien

<400> 70
agaatgatca tcatataggc aatggttctc tagatcatgc tcgagcggcc cagggtgatg 60

gatcgccggg caggtagaga agcctacctg ccctaatggc tcagggctat atcacctccc 120

ggataaccct ggcccttggg actccatcat ctccttgaag tagcactgag aatccaagaa 180

gaggctccgc tgctttttgc acatgttact gagttacatc tcaggaagat tttaagcac 240

gaggaaggaa aatacaggcc tggccaagca gggtcccctt ttcggtatca tctttgttcc 300

taataagcaa tcaaggggt gggtgtgttg gctggtaaag gaactactaa gattcag 357

<210> 71
<211> 1589
<212> DNA
<213> Homo sapien

<400> 71
tggatgtata aggcaggttt tatagaacca tttagattca acattaatgg tagagtggca 60

ttttaccaaa aaaatgggtg tatttgattt gggtgctgta aggcaatttg ctaggacatg 120

ggataatcag attacacaaa atctaggcag tgaacaagtc ttcatcctgg cccaggaaat 180

agcctattaa aaaatgctgt ccaggccagg cacagtggct cacacctgta atcccaggac 240

ttcgggaggc tgaggctgga ggattgcttg agttcaagac cagcctgagc aacacagtga 300

aaacccatct caaaaaaaaa aaaaaaaaca aaccaaaaaa actggtccag ctggtgggac 360

tttcaggatt caggactgct gggggatcaa gcccaaggat gtttttcaga gctctgtgga 420

atttaagatg ctcgaaaaga gtgtccaaga ggagtaaagg tcatgacaga atttactccc 480

aggacaatct aagttctgcc acaagtaccc gtggtgtctg ttcccacaac aatggccctt 540

tcacaagctc ttgcctcaca acccctgcag aagtccttca acaaaactaa taatagacta 600

gtgaaaccta ctcctcacat gggtaagagt tgcagtgggc aggtgaccct cctccctgcc 660

cccatccttt gcttcctcaa gctccctgcc aacctctgga tacatcaatg ggaaggaaac 720

cagggaagca tagacctata gtacaacagg ggtgtagtga ccactggacc tgatgaagcg 780

aatctgcctg aaatttaata cgccttttta tttcccttct gtgttaaatc aaatgatctg 840

ttctgcactg agccaagcaa gttactttta aaactggtgg caccactcat ttgggacttg 900

gagactgctt ttatccagaa cctgttaaga gaacagggga tttaaataca aaggaatatg 960

aaggtggtgt gccttaaaga cagtctaaaa ttaggtttta gtttgttaca ttattttgaa 1020

atattaaaca tgaaaatgtt aaatcaggat gtgtgagttt taagatgttg aacactgtcc 1080

tacatcagtg aggagggagg caataaagta atttcagagt aaaacagatt gaggatgaaa 1140

ccaagacaga agtgatgatt tggcttttta tgatttttgc tgtggaaatg gcagtcgtgt 1200

gacttttcac ttgcagttta aatgaaaggg tgaagagaag cctacctgcc ctaatggctc 1260

agggctatat ccacctcccg gataaccctg gcccttggga ctccatcatc tccttgaagt 1320

agcactgaga atccaagaag aggctccgct gctttttgca catgttactg agttacatct 1380

caggaagatt tttaagcacg aggaaggaaa atacaggcct ggccaagcag ggtccccttt 1440

tcggtatcat ctttgttcct aataagcaat caagggggtg ggtgtgttgg ctggtaaagg 1500

aactactaag attcagaagc ttgtagtctt cattattttg ttttacaggt taaaataaac 1560

cacttgactg ggaaaaaaaa atggcggtc 1589

<210> 72
<211> 471
<212> DNA
<213> Homo sapien

<400> 72
cgtgcggggc tgatagactc tatagtgccc attgtttcct ctagtattca tgctcgagcc 60

ggccgccagt atgataggaa tgcccgggca ggtcggctct cggggaggct gctcagaccg 120

ccgagtccac agcagctaca accgtggtgt tcttgatttt attctgcaga gtgagctctc 180

cacgtttgct ttctggagga cccaggtgac cgctcacctc cccttcctcc tggagccctg 240

aagtcggagg ccctgagcca tggacggtat ctgaggatcg ggtttagcgt atctggccgg 300

agaaattggc aacattgcct acgcaataaa acccaagcgg tttccgaaaa aaaaaaaaaa 360

gagaaaaaaa aaaaggctgg gggtactctt gggccttaag ctggttccct ggggtggcaa 420

tttgtgtttt cccgcctcca catatccccc cacaataaac cggaaagaaa g 471

<210> 73
<211> 772
<212> DNA
<213> Homo sapien

<400> 73
ggatgtatac tcactatggg gcattggtta tctagatcat ctcgagcggc gccagttgta 60

tggatggccg ccccgggcag gtcgtctaac atggcggcgg ctgcggggag agggaagcgt 120

tttactggag ctgcattgtg agcacaaagc gaaagccaga gggggagggc agagaccagg 180

cagccgcccc gactggcctc cttaggcccc cctctaaaaa aaaaaaaaaa ttcgagccac 240

acccacgatt tttttggatt tcaactattt tagtcctttg gtagattcaa actacttgag 300

gttattattc tttcattctt gcaattcagc ctttttgttt ttattgcttt tcattcgcag 360

gacagtgtta aaccttcttt gttttagcag cggggccgtt gttgttattg ttattgttag 420

acgaaaaggt tctgggcaga gttgaattac attaacaatt gacgttaaga tctcagaggt 480

tggaagggga gaaaccaaat ttagtcgttt tgtaaaaacc gaggtaatta cgtctgtgac 540

tattcatgtt accttggaat tttaccttat aaagtaaaat gaagccaaaa aaaaaaaaaa 600

aaagacaaaa aaaaagagtg ctgtggggta atcctggggg cctaaggatg tgttccctgg 660

This page is not part of

the pamphlet!

WO 02-40535

3/3

Date: 23 may 2002

Destination: Agent

Address:

tgtgtcgtac tattgttgtc ccggttccca atttcccccc ctattcttct gtggacgaca 720

aaaggggatg aggccgaaga gcgggggag agcgacggga atggcggaca gg 772

<210> 74
<211> 1061
<212> DNA
<213> Homo sapien

<400> 74
ttggggtctg tccgctcggt taccatgcac tcgagacctg tcgagcgtcc cctcttcttc 60

cgtaggagag aagtgtgttt agaatcttaa ggtagagact gcctttccgg caggcccatt 120

ttgaatgggt cttcgatttg ctaccccgcg gcccatgcga tgggctcccc tgcgtttccc 180

tctttgtttc aattgtttag gtgtcccgcc cgagcctcag gctcagctca atcgcgagat 240

gattttctgc agcgactttt tgttcctagg ggactgtgaa ggggcggggg actgccacga 300

tttagattcg ttgggggctg ggtcctgggg agactggaga ggatggctgg gactcggggc 360

acatggagag agcgtctaac atggcggcgg ctgcggggag agggaagcgc tttactggag 420

ctgcattgtg agcacaaagc gaaagcagag gggagggca gagaccaggc agccgccccg 480

actggcctcc ttaggccccc ctctaaaaaa aaaaaaaaat tcgagccaca cccacgattt 540

ttttggattt caactatttt agtcctttgg tagattcaaa ctacttgagg ttattattct 600

ttcattcttg caattcagcc tttttgtttt tattgctttt cattcgcagg acagtgttaa 660

accttctttg ttttagcagc ggggccgttg ttgttattgt tattgttaga cgaaaaggtt 720

ctgggcagag ttgaattaca ttaacaattg acgttaagat ctcagaggtt ggaaggggag 780

aaaccaaatt tagtcgtttt gtaaaaaccg aggtaattac gtctgtgact attcatgtta 840

ccttggaatt ttaccttata aagtaaaatg aagccaaaaa aaaaaaaaaa aagacaaaaa 900

aaaagagtgc tgtggggtaa tcctgggggc ctaaggatgt gttccctggt gtgtcgtact 960

attgttgtcc cggttcccaa tttccccccc tattcttctg tggacgacaa aaggggatga 1020

ggccgaagag cgggggaga gcgacgggaa tggcggacag g 1061

<210> 75
<211> 426
<212> DNA
<213> Homo sapien

<400> 75
tgtgttttaa gggagctact gtcttaccca aaacctgtga atataacagt gtttttctca 60

tggaattgtg ctcattattc agccagtcgt taatgaattc attcaacaag tgtctctgag 120

aagctagaac actggggatt acagagggaa cacagagaca gatctctggc cctcccagaa 180

actggacagc tattgatggg acagtggtct acatcccatc cagtgtgatc atctcggtga 240

cggactctta cctagaggct gttcctatct gggggtgact ggttgggaaa cctcccaaag 300

tgtgtgcttt tgcgtaaccc gatgttttac ctagggttcc gcgttaacaa gaaaaaaaaa 360

acgtgtgttc tgtctttctg tgatcgcaca gaacacatca caagaagaaa aagaggagga 420

agaaag 426

<210> 76
<211> 977
<212> DNA
<213> Homo sapien

<400> 76
ggatcatcat attagggcga atggtgctct agatgcatgt cgagcggcgc agtgtgatgg 60

atgcgtggtc gcggcgaggt acatactcct ttctggggag agaatgctcc ctaccatata 120

gttgacggtg gttaggaact ctcccttgc cctacctacc ttcctttaa tagcagaatt 180

cctatttttc ccttgattat gtgtattgat caccctgcaa tcctattatg tatctgagtt 240

gtcgtgtgta gtgtgtatag tagtgtgtta tgggggaagg gaggggttcg agtgaagata 300

actgatggtt agtggctttt tcttccatac attagtgacc caccatcggc atgcccagtg 360

ggaccatctg ctcgtggtca tgatgctgca tgtctgtgga ttcatcctgc gagggagggt 420

agtgaagctc aacaatgtcc tttgtttgga gatttttatt tttgcataag tagtccatcc 480

tatacagata gctgattaac tgtattcccc tttcccctat ggctgctggt gtaaataaac 540

tgcatctccc cattggtaaa cagtactaaa atttaaaaa ttttgcagac aaaaaataaa 600

aaagaaatca caaatcttgg gggtaacccg ggggcccaga agctggtccg ggggggaac 660

tggtgccgcc ggccacgatt ccccctaaat ctctggcacg cacaaatgtt tttaaacaca 720

ccgagacatc tcccttacaa gaattttaga caaacaagtt tacagcatat cctaggttac 780

cagaaaaaat ccttggtggc taggaatact ttttaaacag gtatcttgac taccattaag 840

actgcttttt ttttccagac agtttccacg cacctggttc tggttcctta atctcggaaa 900

ctccgaaaaa aaaacaaaca aaacaagtcg gggtacagg caatgcgtcc cgggtgaaat 960

cgtgtatcgg ccaacac 977

<210> 77
<211> 4025
<212> DNA
<213> Homo sapien

<400> 77

cggcacgagg aggcccagga ggctgggtga ggcgctgaga cggtttggcg gtgagtcctg 60
ggccaggcgc agctgaaagg cccgcaaccc gggaaacgtc aaaacaaaca gaaggacttg 120
ggattccgga gcagtcgccc ctatcgctgc tcctgcagtt gcggacgcca ccgaccccgc 180
cgccggagga ctgggcactg aaaggcctct aggcctaggc gcggcccgcg gagccagacg 240
tgttgctgcc gtgagtaaaa cgagcgccct ctccgcactc gtttacaaat taaaatggag 300
gaaatttcgt tggccaacct ggatactaac aagctagagg ccatcgctca ggagatttac 360
gtagacctga tagaggattc ttgtttggga ttctgctttg aggtgcaccg ggcagtcaag 420
tgtggctact tctacctgga gttcgcagag actggtagcg tgaaggattt tggcattcag 480
ccagtggaag acaaaggagc gtgccgcctc ccgctttgct cccttcccgg agaacctggg 540
aatgggcctg atcagcagct ccagcgctca cctccggaat tccagtagct gcaaaatgag 600
agtctgaaag tggccaggac aataacatag actggtcctg tggcttcgag gagtaagcta 660
agtagaaaaa agtagaaaaa tcagacaaaa gttttaattc cccttgaag atcctagcat 720
ttaaaaaccc aaagtggata atttaggaat cctttttta aagtgtatta cctggagcaa 780
gctctgaagc cctgggcagg aggagctgca cagcctgcgg gccatgcagt gcctgttgat 840
ctctaaacac accaggatgt gcgcaagatc ctgtagtgcc cccagtgcac aggtgagcag 900
ttgtgtgccc agcatataaa atttttggtt cctcagcctt tctgtctgcc tgatgtcaag 960
ggcttcctgg acagtttgga cgttacagtt cgtcaggccg tgatcagtgg cctgcagtgg 1020
gactgctcct ttgatatctg aacctctgtt atgggcttct ctgagacaag taaatgtcag 1080
gtgcaagatc tggatactaa cagtttcagt ttgggaaatc caagaaaaag aattatcaag 1140
tttgataggg aagctctgta gccttgactc cagcaagaag aaaaggtcaa aaccacgtgt 1200
ttcccaaaag tccagactac aatgattcag ctgacttgag gacaaggcct agcatttggc 1260
tgagcagagc cctcttcctt gccctccaac ctggtggcat aggcttggca aatggacaac 1320
ttggttgtcc agacaggttg aggattcggt tatgatcccc tggggaggta gcagggacct 1380
ctgcaactat gcatgatttc tcaaacttca agattcatgt ctggatgtat tatgctgtgg 1440
atataagttt agtagggcgg tcatttccta ctctgagtta ctggttacct agccagtcca 1500
tgggtgtgac ttggtcctta agtcaggtca ctatctgcct cccaccctgg gggcaggact 1560
gaagtataga agagcatcat ggctgtgcag gaggctgtgg tttgaaaact gagcccagag 1620
ggcactttca gctgccctca ataatgtgaa tggattagtg ctaggagcca aggagcagga 1680
ctggattatc tcatctgact gtgtgcagaa tcctgttgaa tgtccctgtt ttctttggtt 1740
gggcagtcag agctctgcta tggtgaacat ccagactgtc accactttct gtctgccgct 1800

cgaaagggat agtcctttcc actcggtccc ctttggatct tcttgacaac aggagcagtc 1860
cttttattgt tagaagtcag agaaagacct ccagaatctc ctgactttag ggaatggtat 1920
aggggaagat gggaagtaag agtcacatat caaaactacc ctccacttta ttccctgagc 1980
gagggtttat gaagtataaa ggggtgggag ccccgaggtg agcgggaacg gtgctgcttt 2040
atttgaaatg ttttcttacc tcattctgtg ccccagtagg gggtccagcc tcatctgtct 2100
ggcttggccc tgtgttcctc ctgtcccctg ctccactgcc tatctggtgc cccaggtgct 2160
gcttgccact ccagctgtca cattgaacag tttcaattca gctcttaatg ctcctgcttc 2220
cgaagcctgc ccaatttctt ttttcttggc ctctgttttt tttttttctt tctttttccc 2280
ttgtttttgt agaagactca gaggagaatc tttcttatgg ctccctctgt tgagattgga 2340
attggaagag aacttaattt tttgtattta aaatgcagtg tcatgcctat aagcatttct 2400
cctatatagg actgctttgc tagtgtgccc tcttgctgtg tcttacttca taaggagttg 2460
tatcttccca cctccatttc aatactgccg gttaggacct aagtagaaga gcagtaaagg 2520
ctgattgaca cacaggggga tggagttggt ccttgtccat tctctcaccc ttgctgtgca 2580
tgtatcaatc cttatcccag aaggtactat ttagactgta tagactgatt tagattacat 2640
actttagagg attaaggaaa ccatagagtt tgggccttgg aactgttact gccttgtcct 2700
agagttgtcc tgatcaggct tggggcctag ttacagatta gtcttaaaga attgcattaa 2760
cttaaaaaaa atcaaacctt ggcaagagct aaaataattt ggagatatct ttgcccttga 2820
cttgtagacg acatctaaga ggatgaagaa aggagagtct aagtgagact ctggcctact 2880
tcctaacaat gtcttggaag tgggatgatg gtaaaggaga aaggccacag tccaatccct 2940
ctgccttcag atagggaact caaatcctga aattactgtt ttctttctgg ccttttctcc 3000
tggttagagg aggaagcgga aagtagtttt gagtaatact ttgttcatat taccccccctt 3060
ttgtttttttg tttctggccc ctctaccaat agggcagtag cctcctgccc tggatgggta 3120
taaggtgggc ttggtccaac aggttcccag agggtacata ctcctttctg ggagagaat 3180
gctccctacc atatagttga cagtggttag gaactctccc tttccctacc taccttcctt 3240
ttaatagcag aattcctatt tttcccttga ttatgtgtat tgatcaccct gcaatcctat 3300
tatgtatctg agtgtgtgtg tgtgtgtatg tgtgtgttat gggggaaggg gggggttctt 3360
taaaatttct gtggtttgtg gctttttctt ccatacatta gttcccacca tcgcatgccc 3420
agggaccact gcctggcatt atcgcatgct gggatcatcg ggggagggta gtgaagctca 3480
ccactgtcct ttgttttgga gatttttatt tttgcataag tagtccatcc tatacagata 3540

gctgattaac tgtattcccc tttcccctat ggctgctggt gtaaataaac tgcatctccc 3600

cattggtaaa cagtaataaa attttaaaaa atgaaaaaaa aaaaaaaaaa aaaaatcaca 3660

aatcttgggg gtaacccggg ggcccagaag ctggtccggg gggggaactg gtgccgccgg 3720

ccacgattcc ccctaaatct ctggcacgca caaatgtttt taaacacacc gagacatctc 3780

ccttacaaga attttagaca aacaagttta cagcatatcc taggttacca gaaaaaatcc 3840

ttggtggcta ggaatacttt ttaaacaggt atcttgacta ccattaagac tgcttttttt 3900

ttccagacag tttccacgca cctggttctg gttccttaat ctcggaaact ccgaaaaaaa 3960

aacaaacaaa acaagtcggg ggtacaggca atgcgtcccg ggtgaaatcg tgtatcggcc 4020

aacac 4025

<210> 78
<211> 674
<212> DNA
<213> Homo sapien

<400> 78
gccttttgtg atggatgagg cggccgaggt tcgaccggcg agggaggaag aagcgcgaag 60

agccgttaga tcagtgccgg atgtggtgac ggcgtgggag actgcgggcc cgtagctggg 120

atctgcgagg tgcaagaaag cctttgaggt gataggtgta tgaaatgtca tcataacaga 180

tgtaaccaaa aacttgataa aaggttgtga aaaaactact aggatcacgc ggcatgtatt 240

gagcatatag gttgctgtag atgaatgttc ttagctgtca tgtttaaaaa tacttctgct 300

tcgttacctc aagtgtggca tgcagcattt tggaaggaaa attgaagacg tgttcaagaa 360

aacatgaaca gaagcaaatg atgaaaatga gcattttact tgacgttgat aacatcacaa 420

taaattataa agaaaaaaaa aaaaaaaaag gctgggggat aactcagggc tcaatagcgt 480

gttcccgtgg tgtgtgacaa ttgggctata ctccgcggcc tccacaaatt cccccacgac 540

caacattgcg ggagacacca aaacagaaaa caggaagaag caaaagcaca aaggccaaag 600

cagacaccaa gacaaacgaa gaagaaaaca ggggaacaaa caaaagagaa gacaaaaacg 660

aaaaaaaaga gaaa 674

<210> 79
<211> 1375
<212> DNA
<213> Homo sapien

<400> 79
atggctgatt tcaggcctgg gatagaaaat atagcagatg gacttggggt atggtctaac 60

aaatggcctg tgtcaaaagg acataggagc aaccttgaag ggaccccag tgacaaaaga 120

tgtaagcagg aggggccat aaatcagggc ctggagttcg gtggcatcaa aagagttaga 180

gctaagtctg ggtgtcactg cgtaaagcgg aggccctggg gagtggacgc gttttcacgg 240

aggcatatta agtcgggaaa agacatagaa gcctgtggaa aagcgttaaa gccggtgcac 300

tcagcccccc ttcgcacccg cggagggggcg gggccgcgta ccggaagagg cggggccacc 360

ggagtgccta agagctgtct tccgatgtcg ctcttccttt cccgcgcgac cggcgaggga 420

ggaagaagcg cgaagagccg ttagtcatgc cggtgtggtg gcggcggcgg agactgcggg 480

cccgtagctg ggctctgcga ggtgcaagaa agcctttgag gtgaaggtgt atgaaagtca 540

tcataacaga tgttttccaa aaacttgtag aaggttgtga aaaaactact aggatcacgc 600

ggcatgtatt gagcatatag gttgctgtag atgaatgttc ttagctgtca tgtttaaaaa 660

tacttctgct tcgttacctc aagtgtggca tgcagcattt tggaaggaaa attgaagacg 720

tgttcaagaa aacatgaaca gaagcaaatg atgaaaatga gcattttact tgatgttgat 780

aacatcacaa taaattatgg agaaaaatac aaaaaaaaaa aaaaaaaaaa agaaaaaaaa 840

agagggagga agaaaaaaaa caaaaaaaag aggggggggg agccaacagc caccccggtg 900

gggccaagcg agtggcaaca cgaccaccag accgacatca cgaccgacgg tagcacccaa 960

aggagaggac taaacagcct caccatgcat agcccagtac aaactagaac gcccgagagg 1020

ggcacacaat agccgtacgg gatgacaatt gaaggaaaca cagcgaccga aggaataaca 1080

accggcgaac aaaccacaaa caatgacccc aaggcacaca cacacgcaac ggtggcgccg 1140

gggaccacga gctccaacgc gagcggacgt aaagaaggct acagggaaag tcgaaaaaca 1200

accggccgag gacgaagagg agacgcggag gccaaccggc gcagccccac agcccccgac 1260

ccgcacaggg agacccggac caccagacgg aaagagccga caacccacca gaacaggacg 1320

acgcacgaac acacacgaca ccgacatcac acacaacagc gaacggaaag gcata 1375

<210> 80
<211> 911
<212> DNA
<213> Homo sapien

<400> 80
gaaaaaagac ggggagaatg atactatggc ccgaatggtg cctctagatc atgctcgagc 60

ggcgcagtgt gatggattgg tcgcggcgag gtctgtggga gcctggccta cagtgtggcc 120

ttccacgtcc accggggccc tcagcctcca gtctcagaca gccctcccag ggctggccag 180

ccagaactga tgtcaccatg cccagagccc cagctcccca tactgcagaa ctgatgatgg 240

tcatgggggg cagtggagca ggggcaggag agcaggatga gcaggaatgc aataatcaag 300

atgatccaga atgagaagga agcggaagac aaggctcagt gtgagaccag ggtcagagct 360

cagcaaactt ccacgactgg ctttgaatca gaatcatttt gcttctcagc cacggcccct 420

gggttataca gccttaaatg gccctgccaa tgctggtcac agcatttccc tagtcctgga 480

gactcgggaa ctaaaacaat caattcccct gagcaataaa attatggaca gtgcaaaaaa 540

aaaacaaaaa aaaaaaaaag gctgtggggg taccccgggg gccatacgcg gtcccgggtg 600

tgaactggtg tcccgctcca tccactccga cacacacaca agcagaaaaa aaaaagagga 660

gaacaccagc gaaaagcgaa aaaacaccac aaggagaaag aagaaaccag agaaggcaac 720

aagaaaacag agagaaaaca agagggcgag ggggaaaaga gacgcgcgca agaaaaagca 780

ggaaccgcag gcagagacag agaccagcaa gggcacacaa cgacgaacaa cgaaacgcag 840

ccaagagcag acgaaagcaa gacacaaagc agacgacgaa cgaggcacgc gaaaggagcg 900

caagagagag a 911

<210> 81
<211> 970
<212> DNA
<213> Homo sapien

<400> 81
ggcggccagc gggtgggtga ggccatcttc cccatctacc cgaggccaga ccaaccccgc 60

atgaacccaa aggctcagga tcacgaggac ctgtaccgct actgtggcaa cctggctctg 120

ctccgggcta gcacggaccc cacagcccga cactgtggga gcctggccta cagtgtggcc 180

ttccacgtcc accggggccc tcagcctcca gtctcagaca gccctcccag ggctggccag 240

ccagaactga tgtcaccatg cccagagccc cagctcccca tactgcagaa ctgatgatgg 300

tcatgggggg cagtggagca ggggcaggag agcaggatga gcaggaatgc aataatcaag 360

atgatccaga atgagaagga agcggaagac aaggctcagt gtgagaccag ggtcagagct 420

cagcaaactt ccacgactgg ctttgaatca gaatcatttt gcttctcagc cacggcccct 480

gggttacaca gccttaaatg gccctgccaa tgctggtcac agcatttccct agtcctggag 540

actcgggaac taaaacaatc aattcccctg agcaataaaa ttatggacag tgcaaaaaaa 600

aaacaaaaaa aaaaaaaagg ctgtgggggt accccggggg ccatacgcgg tcccgggtgt 660

gaactggtgt cccgctccat ccactccgac acacacacaa gcagaaaaaa aaaagaggag 720

aacaccagcg aaaagcgaaa aaacaccaca aggagaaaga agaaaccaga gaaggcaaca 780

agaaaacaga gagaaaacaa gagggcgagg gggaaaagag acgcgcgcaa gaaaaagcag 840

gaaccgcagg cagagacaga gaccagcaag ggcacacaac gacgaacaac gaaacgcagc 900

caagagcaga cgaaagcaag acacaaagca gacgacgaac gaggcacgcg aaaggagcgc 960

aagagagaga 970

<210> 82
<211> 681
<212> DNA
<213> Homo sapien

<400> 82
gcactgacca tataggcatg ggtcactaga gcagccgagc ggcgcagtgt gatggatgcg 60

tggtcgcggc gaggtacaga gtcctgttat ttttctcttc ggccctattt ggctgctttt 120

attaatgcat cagaacttta tgttataatc atatggattt atacgtaaat taagaaaaaa 180

tgtccaattt cattcagttc atatgttcta aacgtattgc tgatcattct taaatgagac 240

tccaggttta cattcttaca taaagtgcag ggatcccgaa gttagcccca aagatcccct 300

tgtccttttt cagacttgct caaatgttac cttatcagtg gggcctttcc tgaccacact 360

ttaaaagacc tcaacaccca cccatgggcc ttgtccctcc ttcccggctt catttttttgg 420

catatactta tcaaatgtga acatatgatg catttgcttt atttatcatc gatcttcact 480

cactggcatg taagctctgt gagtgcaaag attttcatct agctatcttc cagaacagtg 540

tctggcacag agaaggagct ctatgaatat gtgttgaatg aatgactatc tttgcctttg 600

taaaccccat gctattggct ctctcttcaa gtggctgaac actgcacccc aagcatgctg 660

gaaggacagg agtccaagcc c 681

<210> 83
<211> 1431
<212> DNA
<213> Homo sapien

<400> 83
agttttgcgt tgcgccttgt tgcttgcgtc cgtcgtttgt ttgcctgtgg cttctgccgt 60

ctttttgtgg gctcgccgtt gcccgtgcct gccgctgtca cccttggcgc ccgcctgggt 120

gctggggtcc gcgattgcag tcctttgata gtgttagtga gggggctgt cgtgcgtgtt 180

gtgtatggtc cgcatggggg agtcattagc atgttgagtt gactgtctcc cggtccgttt 240

aacgtgcgtc tggaaggtac atttttgtaa atcaagtagt tggaactaaa tccaacactg 300

ataattgcca tttcaacact gatctgaaaa gtgaattaga agctgtacaa tatcatcatt 360

agaaattctg catatggcta ataaatattc cttttaaaat taatagagtc taaagtcttc 420

caaatgatct ttacagatag agtgggacac tatagaattc tgattatatg atttagattt 480

tagggatgtt ttaacatttt caaaccacta gaaggacatt gggaacagaa agtaatagag 540

ccaacgtcac gtggtaatga tcaatagtcc agttctacga ggagaacaat tttaagctct 600

tcactgaggc caattctgct gtattctaat tccttttagg ttcttggtgg tagagtaatg 660

agctatgacc atctctggaa tactggtgag gaaaatggca gcagtaaaga aatgaggaaa 720

atattaccta attaatgata aagttaggtc cagtacagag tcctgttatt tttctctttg 780

gccctatttg gctgctttta ttaatgcatc agaactttat gtataatcat atggatttat 840

acgtaaatta agaaaaaatg tccatttcat tcagttcata tgttctaaac gtattgctga 900

tcattcttaa atgagactcc aggtttacat tcttacataa agtgcaggga tcccgaagtt 960

agccccaaag atccccttgc ctttttcaga cttgctcaaa tgttacctta tcagtggggc 1020

ctttcctgac cacactttaa aaacctcaac acccacccat gggccttgtc ctccttcccg 1080

gcttcatttt ttggcatata cttatcaaat gtgaacatat gatgcatttg ctttatttat 1140

catcgatctt cactcactgg catgtaagct ctgtgagtgc aaagattttc atctagctat 1200

cttccagaac agtgtctggc acagagaagg agctctatga atatgtgttg aatgaatgac 1260

tatctttgcc ttgtaaaccc catgctattg gctctctctt caggtggctg accactgcac 1320

cccagggcat gctggaaaga caggagtccc aagccctccc ttctgctcta ctccaagctt 1380

ttcttcttgg gtgcattgac tcaagtcagg tagtacttct ctatgtctga g 1431

<210> 84
<211> 626
<212> DNA
<213> Homo sapien

<400> 84
gcgtggtcgc cggccgaggt tgaggcctcg gttcaatgag ggccccaggc aggcgacggc 60

cacacccagt gtaaacgctg catttctaca acagccacct gtgcaggccc tgcatgctct 120

gtaacctggg gatttggtct tctgaaaagg gcaccagatg aaaaactgct cttaagcctc 180

tgttaactag acacagcagt agaactccaa ggtgttgatc cttggattca tgtttctcaa 240

cttcagatga ccacacatca ctccttcctg accactgggc atccatccca ccaggagctc 300

ctaatctgag agctgttaag aaagtcctcc aaaagtgctg actgcagaag taggtagctt 360

ctgctcaaga tgacagaaca agattaactt ttgtattctt cagcaccttt tttattttcc 420

attatcacac tttgataccc tctaaaacat ttagaacacc ctttctagaa cgaaaaaaaa 480

aaaaaagaaa aaaaaaaaaa aggctgtggg gggtactgtg tggccatagg gtgttcccgt 540

ggggtgaatt gtgttctcgc ccaaattccc cccatttgca caaaaagtga gcgggaaagc 600

acggatccct atatgtgtgg agaaac 626

<210> 85
<211> 779
<212> DNA
<213> Homo sapien

<400> 85
ggatgatacg actcactatg ggcgaatggg cctctagatg ctgctcgagc ggcgcagtgt 60

gatggatgcg tggtcgcggc cggaggtacc catctcaatg agcacatagt aaacgtttaa 120

tacctggtag ctatgggtta ttattaacaa ggtattagac tataagaaca aacgatagga 180

caattcaaat tgttgtgaca gtaaaatatt aaatattttc aaagtggtcc agttaaactc 240

ttgactgaat agtggtttaa gaaacaatgt tagaatgacg tggtttcacg atttaacgag 300

gttaagcaaa tggaaatatc aattaagaat cgtggggtgt ttctctactg agctcagcta 360

gtgctatgcc aagtgaagtg aactaaattc tctggttctt tgtggaaaat cattctgaag 420

tatttgctct aaaaatagct tttggggcct gaattcagcc cttaccccat ctcgcacact 480

tctagtgtcc ccgcagccag aggaccaaga tgattactcg tggggccttg ggcctactta 540

agagactcaa gcttgggtgt tcacaggact gttgacttgt aattctaaca tatagatttc 600

acttaagttc aacagagatt ctattgaagt cacctgcgtc tggcgaaagg gctgttctag 660

acgctggaga tccatcaata gacgaaaaaa aaaaaaaaca aaaaaaaaaa aagcctgggg 720

gacccggggc caaagcggtc ccgggtggaa tttggtttcc cgtcccaatt cccccaatt 779

<210> 86
<211> 462
<212> DNA
<213> Homo sapien

<400> 86
acacgagtgt gtgtgggtat gcatgtgccc atgagagaga gtatgcatgt gtgtgcatac 60

gaacacaagt tgctgtgctg gagaggaagc tgggaaagga gaggagagca tgcactttta 120

gtcatccaca tacattccta tgtgtgcaca cacaacatcc acccagagcc tgtctcccaa 180

atcgatggct caaagtcact ttcttatcgt agaccagacc ccacttagac cagcggcttc 240

aaccttggcc tgcacattaa gatcacctga ggagcttgta aaaatcccaa tgaccaggca 300

acaccctaga ccaatacatc agaatttctg gagatgaaat tgggcatcaa tacttcatat 360

caatatttaa tatttatata atctccttgg gtgattccaa tttcctgcca gcgctgagtg 420

ctcctctgca tagaaagccc ttttcctact cccctgctca ga 462

<210> 87
<211> 911
<212> DNA
<213> Homo sapien

<400> 87
gaaaaccaga accacacacc gggaaaacta gagaccaaaa aactagccta taacaagaac 60

ccaaaacaag accaacccac agaaaagact acaaaaacag aagctgcaca cacacacata 120

aaaaggtgtg cacacagggc acaatgaaaa aaaaaccaga aaaaacaaac ggcccctgaa 180

agggcaccct catccctata aggcctgtaa ccggtgcacc cagagcagac aagacaagga 240

gagtgtgcta caaacatcca caggtgactc tgtgaccaca aacccaaggc tggactgcaa 300

agtgctttca cagggcccca tgagggcagc tcctcgtcat ttatattttg ctgagggtct 360

ccttgaatgg ctgcttgcat aaaagtgttt agaagactgc cgttggaatc tgaatctatc 420

tgaaatgtaa ttccatttcc tggaaatgta cacgagtgtg tgtgggtatg catgtgccca 480

tgagagagag tgtgcatgtg tgtgcatacg aacacaagtt gctgtgctgg agaggaagct 540

gggaaaggag aggagagcat gcacttttag tcatccacat acatacatat gtgtgcacac 600

acacacatcc acccagagcc tgtctcccaa atcgatggct caaagtcact ttcttatcgt 660

agaccagacc ccacttagac cagcggcttc aaccttgcct gcacattaag atcacttgag 720

gagcttgtaa aaatcccaat gaccaggcaa caccctagac caatacatca gaatttctgg 780

agatgaaatt gggcatcaat acttcatatc aatatttaat atttatataa tctccttggg 840

tgattccaat ttcctgccag cgctgagtgc tcctctgcat agaaagccct tttcctactc 900

ccctgctcag a 911

<210> 88
<211> 771
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (740)..(740)
<223> a, c, g or t

<400> 88
cggccgcccg ggcaggtctg cgtgggctct gtgagtgaag ggaggcttca ctactttctg 60

tgagcagtaa ggactggtat ctttctgtga gcaataagga ctggataaag actgcctatc 120

cttgtgtcgt gtcagcacca atacaataag gagggtttta atgtgaagca ggcaatcttc 180

ccaagcccct tctggtcttg gatgaaatag ttgcacagag tattgcacca aaaatacaca 240

atggaggctg aaaagttcaa catattttaa gtcaattaat caaattgcat tgattcttga 300

tgctttctta gaggcctaca tgatttctta gattgctctg ataaactatc ataaggggtc 360

cacttcccct catttagctc ccccagggat ttcttttccc ccatgtcata cacccagtcc 420

taaatcaacc cccaaggcta tccttccatc ccttctgcag agggaacttt tgtcagactc 480

tgcaacaaac tcctagctct atccagagtg tcctctgctg ctaagattgg tatctttctc 540

ctcaaaagcc tggatggtga atgggggtgc attagtcaga attctccaga gaaacagaaa 600

aaataagatt cgcgtgtgtg tgcaacatat attaattaat acaatatatt tattttacaa 660

caacaagacc aaaaaaaaaa aaaagggggg ggaacacctg ggcaaagggg tcccgtggga 720

attgttctcg ccccatcaan cacaaaaaac aaggaaaacg gacccaacaa c 771

<210> 89
<211> 2238
<212> DNA
<213> Homo sapien

<400> 89
cattgctttg ccctggagca gctattttaa gccatctcag attctgtcta aaggggtttt 60

ttgggaagac gttttcttta tcgccctgag aagatctacc ccaggagaat ctgagacatc 120

ttgcctactt ttctttatta gctttctcct cattcatttc ttttatacct ttcctttttg 180

gggagttgtt atgccatgat ttttggtatt tatgtaaaag gattattact aattctattt 240

ctctatgttt attctagtta aggaaatgtt gagggcaagc caccaaatta cctaggctga 300

ggttagagag attggccagc aaaaactaag ctgcctatca gtttgatttg gacaacttga 360

catttatttg agacattaag ctactttctg gtaatatatt aggcatttct gcaatagctc 420

tttcaggtaa ctgaatatta ttaagcatag ttttatcttg ctttgattaa acctcttagg 480

caaaaaatgg aacttcataa gctaatacat tagaaagggg ttatgattat aaatcagaaa 540

tgcttgtgac attaagaaat gaggcacttg tgaaatttct ttgaaatagc cagctcctct 600

aatgtgtctt caaaatataa agtgattcac aaaggcatgc atcacaccta tttgtagcag 660

cccattcatt acataaacca gggcatacct gtgtgggctc tgtgagtgaa gggaggcttc 720

actactttct gtgagcagta aggactggta tctttctgtg agcaataagg actggataaa 780

gactgcatat ccttgtgtcg tgtcagcacc aatacaataa ggagggtttt aatgtgaagc 840

aggcaatctt ccagcccctt ctggtcttgg atgaaatagt tgcacagagt attgcaccaa 900

aaatacacaa tggaggctga aaagttcaac atattttaag tcaattaatc aaattgcatt 960

gattcttgat gctttcttag aggcctacat gatttcttag attgctctga taaactatca 1020

taaggggtcc acttcccctc atttagctcc cccagggatt tcttttcccc catgtcatac 1080

acccagtcct aaatcaaccc ccaaggctat ccttccatcc cttctgcaga gggaactttt 1140

gtcagactct gcaacaaact cctagctcta tccagagtgt cctctgctgc taagattggt 1200

atctttctcc tcaaaagcct ggatggtgaa tgggggtgca ttagtcagaa ttctccagag 1260

aaacagaaaa aataagattc gcgtgtgtgt gcaaacatat atataaataa ataaaaatat 1320

atttatttta aggaattgac tcacatgatt ttgaaggcag gcaagcccaa agtctgcaag 1380

gggtgggcca gcagagagct ggtgctacag tgcaggtctg aaagttgcca gagtcccttt 1440

tatagagaag cttaaaaaat atttgttgaa ttaaatgtct taggtagaga caaattgaat 1500

taaagtggtt acgtaaatac tcttaggtga aatttgtgca aattatgagt ataaagaggg 1560

tgagctacag aactctccat gaccactcaa gaatgggacc caaaggcaaa tgataactta 1620

ttcattcatc aagggataaa cttgagttat caaatggtta atagataaag atcccactct 1680

gccttggatt tagccatcgc agtccctgtg taaaacccca ccggctttca ggcatggaaa 1740

aaaatttcac aaaggccttg aaagaaggca tatgaagacc aaaaggatat tgagcagggt 1800

tccatacttg aataaaggaa gaaactccct tcaatgttta ctaagaaata gttagaatat 1860

agcatatcct gtatagatga aatgtcctag cgaaaatgtt taaatgtatt taaagaagag 1920

ctagattttg tggggggaga gaagggaata ggataggctg gtgggaggaa gaatgaaaaa 1980

aaaaaaactt cgtggaaatt ttggagtaga ttgtgaggag gggtgactca ctttaaagat 2040

gtatcatttg gaggtagaca cgattttcag agcacaaaac agataaggaa agatgtatgc 2100

tgattgtgtg gagcccaagg gaaggcgggt tttggagtct tacgtggctt gtttctgaaa 2160

gggagacatg tgaagcctac tcctgaggcg ggaggctggg acatgacatc agacctataa 2220

tgataacaat agggctgg 2238

<210> 90
<211> 631
<212> DNA
<213> Homo sapien

<400> 90
gccgaggtgg gaggatcgct gaatttcagg agttcaagac cagaccgggc aacaaagtga 60

gaccttgttt ccattaaaaa agaaaagaaa attagccagg tgtggctgta agcaccaggt 120

gtggccgtaa gcactgtggt ccccagctac atgagaggct taaggcaggt gaatctcttg 180

agcccaggag ttctgcagtg aatcatacct gtgccactgc actccagcct ggcagacaga 240

gcaagatcct gtctcacaaa gaaaaagaca aaaaacagtg tataaactaa tccagaaaaa 300

ggaagcataa acagaaatgt aaaagtagaa atagctacag gcagaacaag gaaatggaaa 360

taatggtaag agcactgtct tttactctgt gcgaatccac gagaaaacag ggaccaaaca 420

ggtggctttc tagaaaactc tcagttacca aaatggttcc cagaaacaca gaaaaatcct 480

caggcacaca acactaaggc agattcataa taataaatta ggaaaagcac acagcacacc 540

tcatgtgctc agcaaaggca attcttaggt gaatactcaa accttcaggg aatagatcat 600

tccacattac tcaagtttcc agagagagag a 631

<210> 91
<211> 471
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (397)..(397)
<223> a, c, g or t

<400> 91
tcgcggcgag gtagaagcgc gaagagccgt tagtcatgcc ggtgtggtgg cggcggcgga 60

gactgcgggc ccgtagctgg gctctgcgag catataggtt gctgtagatg aatgttctta 120

gctgtcatgt ttaaaaatac ttctgcttcg tcacctcaag tgtggcatgc agcattttgg 180

aaggaaaatt gaagacgtgt tcaagaaaac atgaacagaa gcaaatgatg aaaatgagca 240

ttttacttga tgttgataac atcacaataa attatggaga aaaatacaaa aaaaaaaaaa 300

aaaaaaaagg cggggcgtag ccagagccat agctggtgcc cggtggtgaa ttggtttacc 360

cgtctccaca attcccacac aaatagcgga agcaacnggc acagcgacaa aggaagcaac 420

tcatgaccga cgcaagtgtg aaaggaacgc gagccagaat acaccacaaa a 471

<210> 92
<211> 1344
<212> DNA
<213> Homo sapien

<400> 92
tcgcggcgag gaagaagcgc gaagagccgt tagtcatgcc ggtgtggtgg cggcggcgga 60

gactgcgggc ccgtagctgg gctctgcgag catataggtt gctgtagatg aatgttctta 120

gctgtcatgt ttaaaaatac ttctgcttcg ttacctcaag tgtggcatgc agcattttgg 180

aaggaaaatt gaagacgtgt tcaagaaaac atgaacagaa gcaaatgatg aaaatgagca 240

ttttacttga tgttgataac atcacaataa attatggaga aaaatacata tttggctaac 300

ttttaattgc tgaacaataa agtgttttct tttaaaaaaa taacaacaga acaaaaaaac 360

tcccgaggaa taagtctcct cctctcctct tcccctcctt ttaaaacatt ggcgcataga 420

aaggcatatg cagggactta taagggtgga aaagacctcc tctttagtga atgtttgtgg 480

ttgcccaagt gaatagaagt gtgtttccca cggtgtgcaa caaaactcta gtgggctaca 540

tagggggga ccttggaatg cacactgtaa agacctgggg ggtcaatgaa acgcttttgg 600

tggcacacgg ccatgtaggg ccactatctc acagaggttg agcgcacgaa atgcgtggga 660

taccacatct aacgcgatct acccaagtgg gtgccgttgt gggaacaccg gtttgtaaag 720

caacagaggg gaactatgaa aaatcacgga gagagatttt tcccaatata taaaccactg 780

cggattaaac gcctataaaa ggctgtaaga acggccccta taaggagagg acacggccag 840

tcagaaccca aaacacgggg ggggctcctt taggacaggc tgcgagacga ccacacacca 900

caagggtgtg gccgaccctc aaacgggaaa gggtagaacc cccaggggag ggtcctcccc 960

aggcccccgg gggaaaacac actacggtgg gacgcatctg agacagagga gactcgaggg 1020

aataaaacgc ctcggcaaaa gagaaacacg tgtggcggtc atagaacgag cccagtcgcc 1080

gacaaattcg atgggtcgtc ggccccggtg gggacacagg agaagaaata ccccagacag 1140

atgaggggtt ttatccaaaa ggcgccatgt gtgcatcatc acgacgtggg acaggggaga 1200

aaggggagtc acccaaaaga gtagggctgc caggtggggc caagtcactg cagaaaggga 1260

cccggggatc tgtgaaattc gcgccacctg ttgcgacgag agagaatgag aagcgggatc 1320

atacggccga cccatgagga acct 1344

<210> 93
<211> 532
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (414)..(414)
<223> a, c, g or t

<400> 93
tcgcggccga ggtgaaccaa gcaacgccaa ttaccaacaa aatccgttgc gccgcagtga 60

gttagctacc ttctatctcc actttgttct gcacgtcgat ctcaggagaa gccagctccc 120

atgttaagaa gttcaaatac ctagagactg cgatggtttg cagaggctca agctaaccac 180

atggaaagac atggagagat attcctgcca accctcaact actccaacta ttctaagaca 240

tcaaacctaa aaacaaaccg caggtcaccc accggtctga agaggaggat gagagacaaa 300

gaaaaaagtg tctggctgcc tctgctgtct acagattgaa gaagatccat ccagctgagc 360

ccagcctaga ccagctgact tctccaataa gcctgtatga aataaatgct tatngttatg 420

tgaaaaaaaa aaaaaaaaaa aggggttggg ggtggccagg gccaaaccgg gcccgggggg 480

aattgggttc ccgctcccca atcccccaca aaaagggaca agggttcggg ga 532

<210> 94
<211> 106
<212> PRT
<213> Homo sapien

<400> 94

Met Ala Cys Asn Leu Ser Tyr Trp Gly Pro Trp Arg Ala Ala Lys Ser
1 5 10 15

Ile Trp Thr Leu Val Glu Val Gly Gly Leu Ala Val Ser Leu Asp Cys
20 25 30

Trp Pro Pro Arg His Ser Lys Pro Gly Ala Ala Glu Gly Arg Leu Leu
35 40 45

Ser Thr Lys Lys Lys Lys Lys Lys Lys Asn Gly Gly Gly Cys Thr Arg
50 55 60

Gly Arg Lys Arg Gly Cys Arg Gly Gly Asn Gly Val Phe Arg Ala Pro
65 70 75 80

Asn Ser Pro His Ile Leu Ala Lys Glu Lys Cys Lys Arg Lys Lys Lys
85 90 95

Arg Lys Arg Lys Arg Lys Glu Lys Arg Lys
100 105

<210> 95
<211> 59
<212> PRT
<213> Homo sapien

<400> 95

Met Val Ala Pro Ile Asp Ala Ala Arg Pro Gln Asp Arg Thr Thr Glu
1 5 10 15

Thr Ser His Gln Arg Thr Asn Thr Val Glu Arg Ala Arg Gln Glu Asp
20 25 30

Gly Gly Arg Val Ser Gly His Thr Ala Asn Arg Ser Thr Cys Arg Ala

```
        35                  40                  45

Asp Gly Ile Gln Ala Asp Pro Gln Gly Gln Gly
    50                  55

<210>  96
<211>  114
<212>  PRT
<213>  Homo sapien

<400>  96

Met Gly Val Phe Thr Phe Val His Pro Gly Leu Asp Ser Phe Leu Arg
1               5                   10                  15

Gly Ser Leu Ala Leu Tyr Ala His Asn Leu Gly Ser Leu Leu Ser Leu
            20                  25                  30

Pro Pro Arg Phe Lys Gln Leu Ser Cys Leu Ser Leu Pro Ser Ser Trp
        35                  40                  45

Glu Tyr Arg Cys Ala Pro Pro Arg Pro Ala Asn Phe Cys Ile Leu Val
    50                  55                  60

Lys Met Gly Phe Leu His Ile Gly Gln Ala Val Leu Lys Leu Leu Thr
65                  70                  75                  80

Ser Gly Asp Leu Thr Ser Ala Ser Gln Ser Ala Gly Ile Tyr Arg His
                85                  90                  95

Glu Pro Pro Arg Pro Gly Pro Thr Ser Ser Ile Tyr Thr Val Arg Gln
            100                 105                 110

Asp Trp

<210>  97
<211>  71
<212>  PRT
<213>  Homo sapien

<400>  97

Met Leu Ser Ser Leu Ala Gln Val Ile Glu Phe Phe Phe Cys Phe Phe
1               5                   10                  15

Leu Arg Gln Ser Leu Ala Leu Leu Pro Arg Leu Glu Cys Ser Gly Ala
            20                  25                  30
```

Asn Ser Ala His Cys Lys Leu Arg Leu Pro Gly Ser Cys His Ser Pro
35 40 45

Val Ser Ala Ser Pro Val Ala Gly Thr Thr Gly Ala Arg His His Thr
50 55 60

Gln Leu Ile Phe Val Phe Tyr
65 70

<210> 98
<211> 62
<212> PRT
<213> Homo sapien

<400> 98

Phe Phe Glu Thr Glu Ser Arg Ser Val Ala Gln Ala Gly Val Gln Trp
1 5 10 15

Cys Glu Leu Gly Ser Leu Gln Ala Pro Pro Pro Gly Phe Met Pro Leu
20 25 30

Ser Cys Leu Ser Leu Pro Ser Ser Trp Asp Tyr Arg Arg Pro Pro Pro
35 40 45

His Pro Ala Asn Phe Cys Ile Leu Leu Glu Met Gly Phe His
50 55 60

<210> 99
<211> 99
<212> PRT
<213> Homo sapien

<400> 99

Met Thr Gly His Arg Thr Arg Pro Ala Tyr Leu Pro Val Lys Ala Ser
1 5 10 15

Ser Pro Gly Arg Tyr Pro Arg Thr Trp Asp Glu Gln Pro Gly Ser Pro
20 25 30

Glu Asp Thr Tyr Leu Ala Arg Arg Thr Ala Ser Ala Ser Trp Thr Ala
35 40 45

Arg Arg Leu Leu Ala Ser Leu Tyr Ser Gln Pro His Arg Gly Pro Glu
50 55 60

Gln Val Pro Gln Gly Gly Thr Ser Ile Ser Ala Leu His Asp Ala Leu
65 70 75 80

Glu Ala Leu His His His Asp Asn Ala Glu Arg Ala Ser His Gly Arg
85 90 95

Pro Gly Lys

<210> 100
<211> 75
<212> PRT
<213> Homo sapien

<400> 100

Met Cys Phe Val Lys Gln Met Leu Glu Gly Ser Met Leu Val Lys Ser
1 5 10 15

His His Gln Ser Leu Ile Ser Ser Asn Gln Gly His Lys His Cys Gly
20 25 30

Arg Pro Gln Gly Pro Leu Pro Arg Lys Thr Arg Asp Leu Cys Ser Leu
35 40 45

Val Tyr Leu Leu Thr Phe Pro Pro Leu Leu Ser His Asp Pro Ala Lys
50 55 60

Tyr Pro Ser Val Arg Asn Thr Gln Gly Ile Ile
65 70 75

<210> 101
<211> 110
<212> PRT
<213> Homo sapien

<400> 101

Met Thr Leu Asn Glu His Ala Ala Phe Lys His Leu Phe Asn Lys Ala
1 5 10 15

His Leu Ala Leu Pro Leu Ile His Leu Thr Leu Ser Gly His Arg Thr
20 25 30

Cys Phe Arg Glu His Arg Val Gly Gly Lys Val Thr Asp Gln Gln Asp
35 40 45

Pro Lys Ala Glu Glu Phe Phe Leu Val Ala Asn Lys Met Lys Ser Leu
50 55 60

Pro Cys Leu Leu Leu Ser Thr Gln Thr Arg Gln Pro Ser Asp Phe Ser
65 70 75 80

Ile Phe Ser Pro Pro Phe Pro Pro Phe Tyr Ser Thr Lys Pro Pro Ser
85 90 95

Ser Ser Trp Pro Val Leu Asn Glu Leu Leu Gly Thr Cys Pro
100 105 110

<210> 102
<211> 61
<212> PRT
<213> Homo sapien

<400> 102

Met Pro Leu His Ser Ser Leu Gly Asn Ile Val Arg Ser Cys Leu Lys
1 5 10 15

Asn Asn Asn Asn Lys Ile Gly Arg Ala Arg Trp Leu Thr Pro Val Ile
20 25 30

Pro Ala Leu Trp Glu Ala Lys Ala Gly Gly Ser Arg Gly Gln Glu Ile
35 40 45

Lys Thr Ile Leu Ala Asn Thr Val Lys Pro His Leu Tyr
50 55 60

<210> 103
<211> 120
<212> PRT
<213> Homo sapien

<400> 103

Phe Phe Leu Cys Phe Phe Phe Leu Glu Trp Ser Leu Ala Val Leu Pro
1 5 10 15

Arg Leu Glu Cys Ser Gly Ala Ile Ser Ala His Cys Lys Leu His Leu
20 25 30

Pro Gly Ser Arg His Ser Pro Ala Ser Ala Ser Leu Val Ala Gly Thr
35 40 45

Thr Gly Ala His His His Thr Arg Ala Lys Phe Phe Val Phe Leu Val

50 55 60

Glu Met Gly Phe His Arg Val Ser Gln Asp Gly Leu Asp Leu Leu Thr
65 70 75 80

Ser Asp Pro Pro Ala Leu Ala Ser Gln Ser Ala Gly Ile Thr Gly Val
85 90 95

Ser His Arg Ala Arg Pro Ile Leu Leu Leu Leu Phe Leu Arg Gln Asp
100 105 110

Leu Thr Met Phe Pro Arg Leu Arg
115 120

<210> 104
<211> 37
<212> PRT
<213> Homo sapien

<400> 104

Met Arg Thr Ser Ser Ser Ile Val Asp Ser Asp His Cys Val Ser Ser
1 5 10 15

Met Ala Leu Pro Pro Ala Val Ser Tyr Phe Ala Pro Ser Gly His Leu
20 25 30

Leu Arg Gln Tyr Asp
35

<210> 105
<211> 67
<212> PRT
<213> Homo sapien

<400> 105

Met Glu Lys Pro His His Ala Leu Ser His Lys Lys Gln Asn Thr His
1 5 10 15

His Asp Asp Thr His Pro Thr Ala Pro His Thr Asn Pro His Gln Ala
20 25 30

Thr Thr Gln His Asn Thr Asn Asn His Thr His His Lys Met Thr Arg
35 40 45

Lys Thr His Thr Glu Gln Thr Asn Thr Ala His Pro Gln Arg Val Ser
50 55 60

Ala Lys Val
65

<210> 106
<211> 164
<212> PRT
<213> Homo sapien

<400> 106

Met Pro Gly Phe Val Leu Phe Phe Arg Phe Leu Leu Val Phe Phe Cys
1 5 10 15

Ser Phe Val Val Ser Cys Ser Phe Leu Phe Phe Phe Arg Val Phe Ser
20 25 30

Phe Trp Arg Ala Val Val Arg Val Phe Ser Phe Cys Phe Ala Phe Ser
35 40 45

Ser Phe Phe Phe Leu Ser Phe Val Cys Leu Ser Leu Cys Cys Phe Phe
50 55 60

Ser Phe Ser Cys Leu Val Ser Cys Val Ala Val Leu Arg Leu Gly Arg
65 70 75 80

Ser Leu Gly Ser Leu Cys Ser Ser Val Ala Leu Phe Pro Pro Val Phe
85 90 95

Phe Phe Leu Cys Ser Pro Val Ala Asp Gly Arg Ile Cys Cys Ala Cys
100 105 110

Leu Ser Phe Phe Phe Phe Pro Leu Phe Leu Ala Leu Leu Ser Val Phe
115 120 125

Val Leu Leu Phe Ser Ala Leu Phe Trp Ser Phe Ser Ala Phe Val Phe
130 135 140

Phe Phe Ile Asp Leu Ser Leu Ser Leu Cys Ala Leu Ser Leu Met His
145 150 155 160

Pro Phe Thr Asn

<210> 107
<211> 82

<212> PRT
<213> Homo sapien

<400> 107

```
Met Ala Trp Leu Gly Leu Arg Gly Leu Thr Phe Leu Pro Ser Tyr Ile
1               5                   10                  15

Asn Lys Lys Asn Lys Thr Asn Ser Val Glu Val Leu Gly Trp Gln Lys
            20                  25                  30

Phe Leu Gly Gly Asp Met Glu Arg Glu Trp Ala Met Phe Leu Arg Ala
        35                  40                  45

Ala Ser Ser Gly Ile Arg Gly Gly Val Gly Thr Phe His Cys Glu Ser
    50                  55                  60

Tyr Pro Lys Leu Gly Ile Arg Asp Gly Leu Gly Gly Ser Arg Asp Leu
65                  70                  75                  80

Gly Arg
```

<210> 108
<211> 1054
<212> PRT
<213> Homo sapien

<400> 108

```
Met Pro Arg Leu Lys Glu Ser Arg Ser His Glu Ser Leu Leu Ser Pro
1               5                   10                  15

Ser Ser Ala Val Glu Ala Leu Asp Leu Ser Met Glu Glu Glu Val Val
            20                  25                  30

Ile Lys Pro Val His Ser Ser Ile Leu Gly Gln Asp Tyr Cys Phe Glu
        35                  40                  45

Val Thr Thr Ser Ser Gly Ser Lys Cys Phe Ser Cys Arg Ser Ala Ala
    50                  55                  60

Glu Arg Asp Lys Trp Met Glu Asn Leu Arg Arg Ala Val His Pro Asn
65                  70                  75                  80

Lys Asp Asn Ser Arg Arg Val Glu His Ile Leu Lys Leu Trp Val Ile
                85                  90                  95
```

Glu Ala Lys Asp Leu Pro Ala Lys Lys Lys Tyr Leu Cys Glu Leu Cys
100 105 110

Leu Asp Asp Val Leu Tyr Ala Arg Thr Thr Gly Lys Leu Lys Thr Asp
115 120 125

Asn Val Phe Trp Gly Glu His Phe Glu Phe His Asn Leu Pro Pro Leu
130 135 140

Arg Thr Val Thr Val His Leu Tyr Arg Glu Thr Asp Lys Lys Lys Lys
145 150 155 160

Lys Glu Arg Asn Ser Tyr Leu Gly Leu Val Ser Leu Pro Ala Ala Ser
165 170 175

Val Ala Gly Arg Gln Phe Val Glu Lys Trp Tyr Pro Val Val Thr Pro
180 185 190

Asn Pro Lys Gly Gly Lys Gly Pro Gly Pro Met Ile Arg Ile Lys Ala
195 200 205

Arg Tyr Gln Thr Ile Thr Ile Leu Pro Met Glu Met Tyr Lys Glu Phe
210 215 220

Ala Glu His Ile Thr Asn His Tyr Leu Gly Leu Cys Ala Ala Leu Glu
225 230 235 240

Pro Ile Leu Ser Ala Lys Thr Lys Glu Glu Met Ala Ser Ala Leu Val
245 250 255

His Ile Leu Gln Ser Thr Gly Lys Val Lys Asp Phe Leu Thr Asp Leu
260 265 270

Met Met Ser Glu Val Asp Arg Cys Gly Asp Asn Glu His Leu Ile Phe
275 280 285

Arg Glu Asn Thr Leu Ala Thr Lys Ala Ile Glu Glu Tyr Leu Lys Leu
290 295 300

Val Gly Gln Lys Tyr Leu Gln Asp Ala Leu Gly Glu Phe Ile Lys Ala
305 310 315 320

Leu Tyr Glu Ser Asp Glu Asn Cys Glu Val Asp Pro Ser Lys Cys Ser
325 330 335

Ala Ala Asp Leu Pro Glu His Gln Gly Asn Leu Lys Met Cys Cys Glu
340 345 350

Leu Ala Phe Cys Lys Ile Ile Asn Ser Tyr Cys Val Phe Pro Arg Glu
355 360 365

Leu Lys Glu Val Phe Ala Ser Trp Arg Gln Glu Cys Ser Ser Arg Gly
370 375 380

Arg Pro Asp Ile Ser Glu Arg Leu Ile Ser Ala Ser Leu Phe Leu Arg
385 390 395 400

Phe Leu Cys Pro Ala Ile Met Ser Pro Ser Leu Phe Asn Leu Leu Gln
405 410 415

Glu Tyr Pro Asp Asp Arg Thr Ala Arg Thr Leu Thr Leu Ile Ala Lys
420 425 430

Val Thr Gln Asn Leu Ala Asn Phe Ala Lys Phe Gly Ser Lys Glu Glu
435 440 445

Tyr Met Ser Phe Met Asn Gln Phe Leu Glu His Glu Trp Thr Asn Met
450 455 460

Gln Arg Phe Leu Leu Glu Ile Ser Asn Pro Glu Thr Leu Ser Asn Thr
465 470 475 480

Ala Gly Phe Glu Gly Tyr Ile Asp Leu Gly Arg Glu Leu Ser Ser Leu
485 490 495

His Ser Leu Leu Trp Glu Ala Val Ser Gln Leu Glu Gln Ser Ile Val
500 505 510

Ser Lys Leu Gly Pro Leu Pro Arg Ile Leu Arg Asp Val His Thr Ala
515 520 525

Leu Ser Thr Pro Gly Ser Gly Gln Leu Pro Gly Thr Asn Asp Leu Ala
530 535 540

Ser Thr Pro Gly Ser Gly Ser Ser Ser Ile Ser Ala Gly Leu Gln Lys
545 550 555 560

Met Val Ile Glu Asn Asp Leu Ser Gly Leu Ile Asp Phe Thr Arg Leu

565 570 575

Pro Ser Pro Thr Pro Glu Asn Lys Asp Leu Phe Phe Val Thr Arg Ser
580 585 590

Ser Gly Val Gln Pro Ser Pro Ala Arg Ser Ser Ser Tyr Ser Glu Ala
595 600 605

Asn Glu Pro Asp Leu Gln Met Ala Asn Gly Gly Lys Ser Leu Ser Met
610 615 620

Val Asp Leu Gln Asp Ala Arg Thr Leu Asp Gly Glu Ala Gly Ser Pro
625 630 635 640

Ala Gly Pro Asp Val Leu Pro Thr Asp Gly Gln Ala Ala Ala Ala Gln
645 650 655

Leu Val Ala Gly Trp Pro Ala Arg Ala Thr Pro Val Asn Leu Ala Gly
660 665 670

Leu Ala Thr Val Arg Arg Ala Gly Gln Thr Pro Thr Thr Pro Gly Thr
675 680 685

Ser Glu Gly Ala Pro Gly Arg Pro Gln Leu Leu Ala Pro Leu Ser Phe
690 695 700

Gln Asn Pro Val Tyr Gln Met Ala Ala Gly Leu Pro Leu Ser Pro Arg
705 710 715 720

Gly Leu Gly Asp Ser Gly Ser Glu Gly His Ser Ser Leu Ser Ser His
725 730 735

Ser Asn Ser Glu Glu Leu Ala Ala Ala Ala Lys Leu Gly Ser Phe Ser
740 745 750

Thr Ala Ala Glu Glu Leu Ala Arg Arg Pro Gly Glu Leu Ala Arg Arg
755 760 765

Gln Met Ser Leu Thr Glu Lys Gly Gly Gln Pro Thr Val Pro Arg Gln
770 775 780

Asn Ser Ala Gly Pro Gln Arg Arg Ile Asp Gln Pro Pro Pro Pro Pro
785 790 795 800

Pro Pro Pro Pro Pro Ala Pro Arg Gly Arg Thr Pro Pro Asn Leu Leu
805 810 815

Ser Thr Leu Gln Tyr Pro Arg Pro Ser Ser Gly Thr Leu Ala Ser Ala
820 825 830

Ser Pro Asp Trp Val Gly Pro Ser Thr Arg Leu Arg Gln Gln Ser Ser
835 840 845

Ser Ser Lys Gly Asp Ser Pro Glu Leu Lys Pro Arg Ala Val His Lys
850 855 860

Gln Gly Pro Ser Pro Val Ser Pro Asn Ala Leu Asp Arg Thr Ala Ala
865 870 875 880

Trp Leu Leu Thr Met Asn Ala Gln Leu Leu Glu Asp Glu Gly Leu Gly
885 890 895

Pro Asp Pro Pro His Arg Asp Arg Leu Arg Ser Lys Asp Glu Leu Ser
900 905 910

Gln Ala Glu Lys Asp Leu Ala Val Leu Gln Asp Lys Leu Arg Ile Ser
915 920 925

Thr Lys Lys Leu Glu Glu Tyr Glu Thr Leu Phe Lys Cys Gln Glu Glu
930 935 940

Thr Thr Gln Lys Leu Val Leu Glu Tyr Gln Ala Arg Leu Glu Glu Gly
945 950 955 960

Glu Glu Arg Leu Arg Arg Gln Gln Glu Asp Lys Asp Ile Gln Met Lys
965 970 975

Gly Ile Ile Ser Arg Leu Met Ser Val Glu Glu Glu Leu Lys Lys Asp
980 985 990

His Ala Glu Met Gln Ala Ala Val Asp Ser Lys Gln Lys Ile Ile Asp
995 1000 1005

Ala Gln Val Tyr Thr Ala Leu Arg Ser Leu Ser His Asp Pro Arg
1010 1015 1020

Ser His Pro His Cys Pro Gln Glu Lys Arg Ile Ala Ser Leu Asp
1025 1030 1035

Ala Ala Asn Ala Arg Leu Met Ser Ala Leu Thr Gln Leu Lys Glu
1040 1045 1050

Arg

<210> 109
<211> 69
<212> PRT
<213> Homo sapien

<400> 109

Met Ser His His Ala Arg Pro His Leu Phe Phe Ile Arg Ser Ser Val
1 5 10 15

Gly Arg His Leu His Cys Phe Gln Ile Leu Ala Ile Val Asn Ser Ala
20 25 30

Ala Ile Asn Ile Arg Val Gln Thr Ser Leu Pro His Leu Ile Ser Phe
35 40 45

Leu Leu Gly Ile Tyr Leu Ala Val Glu Leu Leu Asp His Met Val Ala
50 55 60

Leu Phe Leu Val Phe
65

<210> 110
<211> 204
<212> PRT
<213> Homo sapien

<400> 110

Met Phe Arg Gly Gly Glu Leu Trp Gly Ala Arg Gly Glu Ile Thr His
1 5 10 15

Phe Leu Thr Thr Pro His Gly Gly Lys Thr Pro Ile Leu Ala Pro Pro
20 25 30

Arg Cys Val Tyr Pro Pro Thr Pro Arg Ala Leu Val Phe Val Phe Phe
35 40 45

Ser Phe Tyr Phe Phe Phe Pro Ser Val Ser Val Cys Ser Pro Trp Leu
50 55 60

Leu Pro Tyr Cys Phe Ala Ser Arg Gly Lys Ser His Ser Arg Lys Asn
65 70 75 80

Gly Ile Tyr Thr Glu Thr Cys Phe Gln Pro Thr Lys Glu Val Asn Pro
85 90 95

Leu Glu Leu Pro Asn Ala Asn Pro Ile Phe Pro Ala Pro Lys Met Thr
100 105 110

Phe Met Glu Arg Thr Arg Glu Glu Thr Lys Arg Ser Lys Arg Gly Phe
115 120 125

Phe Tyr Thr Ala Ser Asp Gly Thr Pro Ser Val Tyr Ala Pro Gly Ala
130 135 140

Arg Ala Pro Pro Glu Leu Leu Leu Thr Phe Ile Arg Ala Gly Met Gln
145 150 155 160

Leu Ala Ser Phe Asn Gln Ser Trp Met Asp Arg Thr Pro Ile Leu Thr
165 170 175

Val Arg Ala Cys Lys Asp Ser Ser Gln Glu Leu Leu Tyr Ala Val Val
180 185 190

Ser Val Gln Ser Arg Asn Ala Ile Cys Arg Glu Val
195 200

<210> 111
<211> 35
<212> PRT
<213> Homo sapien

<400> 111

Met Leu Thr His Thr Phe Ser Arg Glu Asn Leu Gly Tyr Val Gln Tyr
1 5 10 15

Met Tyr Phe Lys Thr Glu Gly Ser Met Ser Phe Leu Arg Asp Cys His
20 25 30

Gln His Gly
35

<210> 112
<211> 99
<212> PRT
<213> Homo sapien

<400> 112

Met Glu His Thr Ile Arg Phe Tyr Thr Glu Thr Phe His Cys Pro Gly
1 5 10 15

Thr Gly Arg Arg Gln Met Pro Ser Ser Cys Leu Asn Cys Lys Glu Ala
20 25 30

Phe Leu Leu Leu Thr Leu Ile Leu Leu Ser Thr Asp Pro Leu Arg Val
35 40 45

Ser Gly Trp Gly Asp Gly Gln Val Phe Pro Phe Pro Arg Gly His Ile
50 55 60

Ser Asp Tyr His Met Gly Arg Asn Leu Gly Gln Tyr Leu Ala Phe Leu
65 70 75 80

Gly Arg Gly Pro Cys Ser Leu Pro Gln Cys Leu Cys Pro Gly Tyr Leu
85 90 95

Pro Gly Arg

<210> 113
<211> 93
<212> PRT
<213> Homo sapien

<400> 113

Met Gly Leu Gly Val Ile Gln Thr Thr Arg Asn Asn Lys Thr Lys Lys
1 5 10 15

Lys Asn Lys Glu Gly Ser Trp Gly Gly Pro Lys Gly Pro Lys Arg Gly
20 25 30

Val Pro Arg Gly Trp Glu Lys Glu Glu Arg Arg Gly Gly Glu Lys Asn
35 40 45

Ser Pro Pro Lys Ile Arg Gly Gly His Asn Arg His Met Trp Ile Arg
50 55 60

Glu Asn Lys Arg Lys Glu Lys Arg Arg Gly Glu Thr Arg Asn Lys Lys
65 70 75 80

Glu Glu Arg Lys Lys Ala Lys Lys Gln Arg Lys Glu Lys

85 90

<210> 114
<211> 69
<212> PRT
<213> Homo sapien

<400> 114

Met Ser Gln Glu Lys Asp Phe His Lys Val Met Ser Ser Leu Lys Ala
1 5 10 15

Arg Thr Gly His Leu His Phe Phe Cys Gly Gly Arg Ser Ser Val Lys
20 25 30

Val Gly Gln Ser Ile Phe Thr Ser Phe Val Ile Leu Gln Leu Leu Gln
35 40 45

Ala Ile Trp Ala Tyr Thr Cys Lys Ser Gln Gly Met Arg Trp Leu Gly
50 55 60

Leu Gly Ser Glu Ala
65

<210> 115
<211> 843
<212> PRT
<213> Homo sapien

<400> 115

Val Asn Asn Glu Ile Lys Thr Glu Ile Lys Lys Phe Phe Glu Thr Ser
1 5 10 15

Glu Asn Lys Asp Thr Thr Tyr Gln Asn Leu Trp Asp Ala Phe Lys Ala
20 25 30

Val Cys Arg Gly Lys Phe Ile Ala Leu Asn Ala His Lys Arg Lys Gln
35 40 45

Glu Arg Ser Lys Ile Asp Ile Leu Thr Ser Gln Leu Lys Glu Leu Glu
50 55 60

Lys Gln Glu Gln Thr His Ser Lys Ala Ser Arg Arg Gln Glu Ile Thr
65 70 75 80

Glu Ile Arg Ala Glu Leu Lys Glu Ile Glu Thr Gln Lys Thr Leu Gln
85 90 95

Lys Ile Asn Glu Ser Arg Ser Trp Phe Phe Glu Arg Ile Asn Lys Ile
100 105 110

Asp Arg Pro Leu Ala Arg Leu Ile Lys Lys Lys Arg Gln Lys Asn Gln
115 120 125

Ile Asp Ala Ile Lys Asn Asp Lys Gly Asp Ile Thr Thr Asp Pro Thr
130 135 140

Glu Ile Gln Thr Thr Ile Arg Glu Tyr Tyr Lys His Leu Tyr Ala Asn
145 150 155 160

Lys Leu Glu Asn Leu Glu Glu Met Asp Lys Phe Leu Asp Thr Tyr Thr
165 170 175

Leu Pro Arg Leu Asn Gln Glu Glu Ala Glu Ser Leu Asn Arg Pro Ile
180 185 190

Thr Gly Ser Glu Ile Val Ala Ile Ile Asn Ser Leu Pro Thr Lys Lys
195 200 205

Ser Pro Gly Pro Asp Gly Phe Thr Ala Glu Phe Tyr Gln Arg Tyr Lys
210 215 220

Glu Glu Leu Val Pro Phe Leu Leu Lys Leu Phe Gln Ser Ile Glu Lys
225 230 235 240

Glu Gly Ile Leu Pro Asn Ser Phe Tyr Glu Ala Ser Ile Ile Leu Ile
245 250 255

Pro Lys Leu Gly Arg Asp Thr Thr Lys Lys Glu Asn Phe Arg Pro Ile
260 265 270

Ser Leu Met Asn Thr Asp Ala Lys Ile Leu Asn Lys Ile Leu Thr Asn
275 280 285

Arg Ile Gln Gln His Ile Lys Lys Leu Ile His His Asp Gln Val Gly
290 295 300

Phe Ile Pro Gly Met Gln Gly Trp Phe Asn Ile Cys Lys Ser Ile Asn
305 310 315 320

Val Ile Gln Tyr Ile Asn Arg Ala Lys Asp Lys Asn His Met Ile Ile

325 330 335

Ser Ile Asp Ala Glu Lys Ala Phe Asp Lys Ile Gln Gln Pro Phe Met
340 345 350

Leu Lys Thr Leu Asn Lys Leu Gly Ile Asp Gly Thr Tyr Phe Lys Ile
355 360 365

Ile Arg Ala Ile Tyr Asp Lys Pro Thr Ala Asn Ile Ile Leu Asn Gly
370 375 380

Gln Lys Leu Glu Ala Phe Pro Leu Lys Thr Gly Thr Arg Gln Gly Cys
385 390 395 400

Pro Leu Ser Pro Val Leu Phe Asn Val Val Leu Glu Val Leu Ala Arg
405 410 415

Ala Ile Arg Gln Glu Lys Glu Ile Lys Gly Ile Gln Ile Gly Lys Glu
420 425 430

Glu Val Lys Leu Ser Leu Phe Ala Asp Asp Met Ile Val Tyr Leu Glu
435 440 445

Asn Pro Ile Val Ser Ala Gln Asn Leu Leu Lys Leu Ile Ser Asn Phe
450 455 460

Ser Lys Val Ser Gly Tyr Lys Ile Asn Val Gln Lys Ser Gln Arg Ile
465 470 475 480

Lys Tyr Leu Gly Ile Gln Leu Thr Arg Asp Val Lys Asp Leu Phe Lys
485 490 495

Lys Asn Tyr Lys Pro Leu Leu Lys Glu Ile Lys Glu Asp Thr Asn Lys
500 505 510

Trp Lys Asn Ile Pro Cys Ser Trp Ile Gly Arg Ile Asn Ile Met Lys
515 520 525

Met Ala Ile Leu Pro Arg Val Ile Tyr Arg Phe Asn Ala Ile Pro Ile
530 535 540

Lys Leu Pro Met Pro Phe Phe Thr Glu Leu Glu Lys Thr Thr Leu Lys
545 550 555 560

Phe Ile Trp Asn Glu Lys Thr Ala Arg Ile Ala Lys Leu Ile Leu Ser
565 570 575

Gln Lys Asn Lys Ala Gly Gly Ile Thr Leu Pro Asp Phe Lys Leu Tyr
580 585 590

Tyr Lys Pro Thr Val Thr Lys Thr Ala Trp Tyr Trp Tyr Gln Asn Arg
595 600 605

Asp Ile Asp Gln Trp Asn Arg Thr Glu Pro Ser Glu Ile Thr Pro His
610 615 620

Thr Tyr Asn Tyr Arg Ile Phe Asp Lys Pro Glu Lys Asn Lys Gln Trp
625 630 635 640

Gly Lys Asp Ser Leu Phe Asn Lys Trp Cys Trp Glu Asn Trp Leu Ala
645 650 655

Ile Cys Arg Lys Leu Lys Leu Asp Pro Phe Leu Thr Pro Ser Thr Lys
660 665 670

Ile Asn Ser Arg Trp Ile Lys Asp Leu Asn Val Arg Pro Lys Thr Ile
675 680 685

Lys Thr Leu Glu Glu Asn Leu Gly Ile Thr Ile Gln Asp Ile Gly Met
690 695 700

Gly Lys Asp Phe Met Ser Lys Thr Pro Lys Ala Met Ala Thr Lys Ala
705 710 715 720

Lys Ile Asp Lys Trp Asp Leu Ile Lys Leu Lys Ser Phe Cys Thr Ala
725 730 735

Lys Glu Thr Thr Ile Arg Val Asn Arg Gln Pro Thr Lys Trp Glu Lys
740 745 750

Ile Phe Ala Thr Tyr Ser Ser Asp Lys Gly Leu Ile Ser Arg Ile Tyr
755 760 765

Asn Glu Leu Lys His Ile Tyr Lys Lys Lys Thr Asn Ser Pro Ile Lys
770 775 780

Lys Trp Met Lys Asp Met Asn Arg His Phe Ser Lys Glu Asp Ile Tyr
785 790 795 800

Ala Ala Lys Lys His Met Lys Lys Cys Ser Ser Ser Leu Ala Ile Arg
805 810 815

Glu Met Gln Ile Lys Thr Thr Met Arg Tyr His Leu Thr Pro Val Arg
820 825 830

Met Ala Ile Ile Lys Lys Ser Gly Ser Asn Arg
835 840

<210> 116
<211> 93
<212> PRT
<213> Homo sapien

<400> 116

Met Leu Ala Arg Met Val Ser Ile Ser Glu Pro Cys Asp Pro Pro Gln
1 5 10 15

Leu Gly Leu Pro Lys Cys Trp Asp His Lys Cys Lys Pro Leu Arg Pro
20 25 30

Ala Leu Phe Ser Leu Gly Ile Tyr Pro Glu Val Glu Leu Leu Val His
35 40 45

Leu Ala Asn Ser Ser Phe Asn Phe Leu Arg Thr Glu His Cys Pro Gln
50 55 60

Trp Leu Tyr Thr Phe His Phe Pro Thr Asp Ser Ile Gln Glu Phe Pro
65 70 75 80

Ile Glu Ser Thr Phe Phe Gln Thr Tyr Phe Leu Phe Phe
85 90

<210> 117
<211> 62
<212> PRT
<213> Homo sapien

<400> 117

Gly Ala Val Ala Tyr Thr Cys Asp Pro Ser Thr Leu Gly Gly Gln Val
1 5 10 15

Gly Ala Asp His Lys Val Arg Arg Ser Arg Pro Ser Trp Pro Thr Trp
20 25 30

Ala Asn Pro Val Ser Thr Lys Ile Glu Lys Ile Ser Trp Ala Trp Trp
35 40 45

Leu Ala Pro Val Ile Pro Ala Arg Leu Thr Val Lys Ala Ala
50 55 60

<210> 118
<211> 53
<212> PRT
<213> Homo sapien

<400> 118

Met Lys Ser Leu Pro Cys Leu Leu His Phe His Thr Asp Thr Ala Thr
1 5 10 15

Ile Arg Phe Leu Asn Leu Phe Pro Thr Val Ser Arg Leu Ser Ile Pro
20 25 30

Gln Ser Arg His Arg His Pro Gly Pro Phe Ser Met Ser Cys Trp Val
35 40 45

Pro Ala Arg Ala Ala
50

<210> 119
<211> 112
<212> PRT
<213> Homo sapien

<400> 119

Leu Ser Glu His Ala Ala Leu Lys His Leu Phe Asn Lys Ala His His
1 5 10 15

Cys Thr Cys Pro Leu Ile His Leu Thr Leu Ser Gly His Thr Thr Cys
20 25 30

Phe Arg Glu His Arg Val Arg Gly Lys Val Thr Asp Gln Gln Asp Pro
35 40 45

Lys Ala Glu Glu Phe Phe Leu Val Ala Asn Lys Met Lys Ser Leu Pro
50 55 60

Cys Leu Phe Ile Ser Thr Gln Thr Arg Gln Pro Ser Asp Phe Ser Ile
65 70 75 80

Phe Ser Pro Pro Phe Pro Pro Phe Tyr Ser Thr Lys Pro Pro Ser Ser

85 90 95

Ser Trp Pro Val Leu Asn Glu Leu Leu Gly Thr Cys Pro Gly Gly Arg
100 105 110

<210> 120
<211> 209
<212> PRT
<213> Homo sapien

<400> 120

Met Gly Arg Trp Glu Glu Ser Gln Ser Thr Gly Gln Gly Glu Asp Ser
1 5 10 15

Gly Ser His Gly Val Ser Pro Thr Ala Ser Ala Pro Leu Cys Cys Trp
20 25 30

Arg Gly Pro Glu Pro His Tyr Ser Leu Tyr Arg Gly Pro Arg Arg Gly
35 40 45

Ala Leu Gly Arg Ser Arg Gly Trp Leu Thr Arg Glu Asp Thr Lys Val
50 55 60

Glu Gly Gly Phe Leu Leu Arg Glu Arg Pro Glu Asn Asn Gln Gly Thr
65 70 75 80

Pro Gln His Ala Val Pro Thr Leu Asp Gly Arg Pro Pro Ser Thr Thr
85 90 95

Asp Asp Ser Gly Arg Arg Ile Gly His Pro Arg Arg Ile His Trp Pro
100 105 110

Ser Thr Leu Arg Asp Cys Pro Met Val Asn Gln Arg Lys Gly Arg Thr
115 120 125

Gly Arg Gly Gln Thr Pro Gly Cys Ser Thr His Gly Thr Thr Phe Pro
130 135 140

Leu Thr Ser Ile Pro Lys Ser Ser Pro Cys Gln Met Leu Ala Ser Ala
145 150 155 160

Asn Val Ser Glu Ala His Met Val Ser Ser Leu Ser Arg Thr Pro Met
165 170 175

Leu Ser Leu Pro Ala Arg Leu Cys Ala Ser Met Gly Asp Asp Leu Ser

180 185 190

Pro Thr Leu Arg Pro Glu Ala Ile His Ser His Asn Ala Pro Ala Arg
195 200 205

Ala

<210> 121
<211> 118
<212> PRT
<213> Homo sapien

<400> 121

Met Asp Glu Arg Arg Pro Gly Arg Tyr Leu Gly Leu Pro Glu Tyr Thr
1 5 10 15

Lys Phe Arg Glu Pro Thr Phe Thr Pro Asp Cys Ala Trp Ser Lys Pro
20 25 30

Glu Ser Ser Leu Pro Arg Gly Leu Phe Gln Pro Ile Pro Leu Phe Trp
35 40 45

Lys Val Ile Leu Gly Ile Glu Thr Glu Asn Trp Asp Lys Gly Ser Leu
50 55 60

Arg Lys Thr Lys Thr Asn Asn Glu Thr Gly Asp Met Leu Phe Ser Leu
65 70 75 80

Asn Pro Ser Gln Ile Cys Cys Leu Ala Leu Thr His Val Glu Ile Cys
85 90 95

Lys Leu Cys Gln Asp Phe Pro Val His Gly Gly Glu Ser His Val Gly
100 105 110

Lys Lys Lys Phe Thr Val
115

<210> 122
<211> 42
<212> PRT
<213> Homo sapien

<400> 122

Met Ala Thr Pro Pro Ala Lys Cys Leu Ser Gln Asp Leu Asp Ser Ser
1 5 10 15

Pro Trp Asp Pro His Ala Arg Glu Ala Asp Cys Ser Ala Pro Thr Gly
20 25 30

Ser Leu His Glu Val Val Pro Gln His Cys
35 40

<210> 123
<211> 59
<212> PRT
<213> Homo sapien

<400> 123

Met Thr Phe Gly Val Pro Asn Ser Val Ser Thr Leu Thr Ser Lys Lys
1 5 10 15

Lys Lys Arg Lys Lys Lys Lys Gly Arg Gly Val Pro Trp Ala Arg Arg
20 25 30

Val Pro Val Val Glu Leu Phe Phe Pro Ser Gln Phe Pro Pro Phe Phe
35 40 45

Thr Thr Met Val Ser Leu Val Lys Arg Glu Lys
50 55

<210> 124
<211> 127
<212> PRT
<213> Homo sapien

<400> 124

Met Gly Glu Leu Cys Ser Arg Met Leu Leu Glu Arg Arg His Cys Asp
1 5 10 15

Gly Cys Val Val Ala Ala Arg Leu Cys Val Lys Arg Glu Ala Glu Gly
20 25 30

Asp Val Ser Pro Asp Ile Ser Lys Val Trp Val Gly Pro Leu Val Pro
35 40 45

Glu Ile Leu Leu Gly Gly Met Gly Pro Ala Leu Ser Gly Thr Lys Ile
50 55 60

Arg Ala Arg Lys Arg Cys Pro Ser Pro Ile Leu Ser Ile Leu Phe Met
65 70 75 80

Ala Glu Lys Ile Ser Ala Gly Cys Gln His Val Pro Met Pro Val Glu
85 90 95

Asp Met Pro Thr Ser Pro Leu Pro Arg Glu Gln Asp Leu Gly Leu Gly
100 105 110

Gln Val Glu Lys Ile Pro Asp Phe Phe Arg His Cys Ile Leu Phe
115 120 125

<210> 125
<211> 121
<212> PRT
<213> Homo sapien

<400> 125

Met Val Arg Ile Leu Ala Asn Gly Glu Ile Val Gln Asp Asp Asp Pro
1 5 10 15

Arg Val Arg Thr Thr Thr Gln Pro Pro Arg Gly Ser Ile Pro Arg Gln
20 25 30

Ser Phe Phe Asn Arg Gly His Gly Ala Pro Pro Gly Gly Pro Gly Pro
35 40 45

Arg Gln Gln Gln Ala Gly Ala Arg Leu Gly Ala Ala Gln Ser Pro Phe
50 55 60

Asn Asp Leu Asn Arg Gln Leu Val Asn Met Gly Phe Pro Gln Trp His
65 70 75 80

Leu Gly Asn His Ala Val Glu Pro Val Thr Ser Ile Leu Leu Leu Phe
85 90 95

Leu Leu Met Met Leu Gly Val Arg Gly Leu Leu Leu Val Gly Leu Val
100 105 110

Tyr Leu Val Ser His Leu Ser Gln Arg
115 120

<210> 126
<211> 67
<212> PRT
<213> Homo sapien

<400> 126

Met Asp Pro Ala Arg Ala Gly Thr Arg Gly Gly Val Pro Ala Pro Pro
1   5   10   15

Ala His Gly Gly Gly Arg Leu Gly Pro Ala Arg Gly Ala Cys Cys Ser
            20   25   30

Pro Ser Arg Pro Pro Arg Pro Pro His Arg His His Ala Pro Val Pro
        35   40   45

Ala Trp Ile Tyr Thr Trp Ala Ser Val Cys Trp Lys Cys Thr Leu Ala
    50   55   60

Gln Thr Leu
65

<210> 127
<211> 64
<212> PRT
<213> Homo sapien

<400> 127

Met Leu Pro Arg Leu Val Ser Asn Cys Leu Cys Val Lys Gln Ser Val
1   5   10   15

His Leu Arg Pro Ser Ala Asn Cys Arg Asp His Arg His Glu Pro Pro
            20   25   30

Leu Pro Ala Thr Met His Ser Glu Arg Ser Arg Asn Arg Glu Cys His
        35   40   45

Ser Thr Thr His Leu Ile Ile Pro Thr Met Thr His Val Ser Gln Arg
    50   55   60

<210> 128
<211> 41
<212> PRT
<213> Homo sapien

<400> 128

Met Asn Phe Gly Lys Ser Ile Met Leu Gln Gly Gln Ala His Ala Pro
1   5   10   15

Gln Tyr Ser Pro Thr Ala Ala Gln Trp Asp Ile Ser Leu Trp Trp His
            20   25   30

Ile Thr Arg Arg Pro Ser Val Leu Ser

35 40

<210> 129
<211> 46
<212> PRT
<213> Homo sapien

<400> 129

Leu Ser Leu Glu His Asp Ala Phe Thr Glu Val His Val Thr Cys Ala
1 5 10 15

Lys Leu Phe Pro Pro Ile Cys Asp Tyr Gly Pro Met Glu Leu Gly Gln
20 25 30

Ser Leu Trp Glu Ala Glu Gly Lys Asp Pro Gly His Phe Arg
35 40 45

<210> 130
<211> 58
<212> PRT
<213> Homo sapien

<400> 130

Met Lys Asp Lys Gly Leu Arg His Thr Glu Thr Gly Gln Thr Asn Gly
1 5 10 15

Lys Pro Thr Arg Pro Ala His Asp Gln Asn Ile Thr Gly Arg Pro Pro
20 25 30

Ala Asn Ala Glu Ala Ser Glu Ser Thr Val Gly Gly Trp Asn Gln Ala
35 40 45

Pro Gly Asn Leu Ser Ala Ala Phe Arg Leu
50 55

<210> 131
<211> 87
<212> PRT
<213> Homo sapien

<400> 131

Met Phe Ser Thr Ser Ser Gln Val Cys Ala Leu Cys Pro Phe Ser Gly
1 5 10 15

Ser Leu Glu Leu Pro Pro Ser Leu His Pro Asp Ser Phe Ala Ile Met
20 25 30

Cys Leu Ile Ser Cys Glu Phe Thr Gly Glu Ala Ile Ser Gln Ile Asn
35 40 45

Gly Cys Lys Cys Ser Lys Lys Lys Lys Thr Lys Lys Lys Ala Gly Gly
50 55 60

Asn Arg Gly Gln Ser Leu Ser Pro Gly Gly His Cys Phe Pro Pro Gln
65 70 75 80

Phe Asn Pro His Lys Pro Pro
85

<210> 132
<211> 264
<212> PRT
<213> Homo sapien

<400> 132

Met Arg Pro Leu Leu Gly Leu Leu Leu Val Phe Ala Gly Cys Thr Phe
1 5 10 15

Ala Leu Tyr Leu Leu Ser Thr Arg Leu Pro Arg Gly Arg Arg Leu Gly
20 25 30

Ser Thr Glu Glu Ala Gly Gly Arg Ser Leu Trp Phe Pro Ser Asp Leu
35 40 45

Ala Glu Leu Arg Glu Leu Ser Glu Val Leu Arg Glu Tyr Arg Lys Glu
50 55 60

His Gln Ala Tyr Val Phe Leu Leu Phe Cys Gly Ala Tyr Leu Tyr Lys
65 70 75 80

Gln Gly Phe Ala Ile Pro Gly Ser Ser Phe Leu Asn Val Leu Ala Gly
85 90 95

Ala Leu Phe Gly Pro Trp Leu Gly Leu Leu Leu Cys Cys Val Leu Thr
100 105 110

Ser Val Gly Ala Thr Cys Cys Tyr Leu Leu Ser Ser Ile Phe Gly Lys
115 120 125

Gln Leu Val Val Ser Tyr Phe Pro Asp Lys Val Ala Leu Leu Gln Arg
130 135 140

Lys Val Glu Glu Asn Arg Asn Ser Leu Phe Phe Phe Leu Leu Phe Leu
145 150 155 160

Arg Leu Phe Pro Met Thr Pro Asn Trp Phe Leu Asn Leu Ser Ala Pro
165 170 175

Ile Leu Asn Ile Pro Ile Val Gln Phe Phe Phe Ser Val Leu Ile Gly
180 185 190

Leu Ile Pro Tyr Asn Phe Ile Cys Val Gln Thr Gly Ser Ile Leu Ser
195 200 205

Thr Leu Thr Ser Leu Asp Ala Leu Phe Ser Trp Asp Thr Val Phe Lys
210 215 220

Leu Leu Ala Ile Ala Met Val Ala Leu Ile Pro Gly Thr Leu Ile Lys
225 230 235 240

Lys Phe Ser Gln Lys His Leu Gln Leu Asn Glu Thr Ser Thr Ala Asn
245 250 255

His Ile His Ser Arg Lys Asp Thr
260

<210> 133
<211> 35
<212> PRT
<213> Homo sapien

<400> 133

Met Thr Ser Ile Ile Arg Ser Glu Thr Arg Leu Ser Phe Trp Met Leu
1 5 10 15

Ser Gly Leu Cys Val Arg Glu Tyr Phe Lys Thr Ala Ser Tyr Val Leu
20 25 30

Leu Gly Asn
35

<210> 134
<211> 39
<212> PRT
<213> Homo sapien

<400> 134

Met Leu Gly Lys Ala Trp Arg Gly Ile Leu Val Gly Glu Lys Gln Ile
1   5   10   15

Arg Cys Leu Leu Phe Cys Ser Val Ser Lys Ser Pro Lys Lys Cys Gly
            20   25   30

Arg Val Leu Leu Glu Arg Lys
        35

<210> 135
<211> 91
<212> PRT
<213> Homo sapien

<400> 135

Met Phe Cys Val Phe Leu Lys Ser Glu Cys Val Phe Tyr His Cys Ser
1   5   10   15

Val Asn Ala Asn Trp Val Lys Phe Val Asp Ser Gln Ile Tyr Ile Leu
            20   25   30

Thr His Leu Phe Val Pro Phe Phe Leu Ser Val Ile Glu Gln Glu Val
        35   40   45

Leu Lys Ser Pro Ile Thr Ser Ile Ser Leu Thr Leu Pro Phe Phe Ser
    50   55   60

Leu Trp Ile Leu Asn Phe Ser Ile Tyr Phe Val Tyr Phe Glu Gly His
65   70   75   80

Ile His Leu Leu Ser Ser Cys Ile Leu Met Asn
                85   90

<210> 136
<211> 38
<212> PRT
<213> Homo sapien

<400> 136

Gln Pro Gly Gln His Gly Glu Thr Pro Ser Pro Pro Lys Asp Ala Lys
1   5   10   15

Thr Ser Gln Ala Trp Arg Arg Ala Pro Ala Val Pro Gly Thr Arg Gln
            20   25   30

Ala Glu Ala Gly Glu Ser

35

<210> 137
<211> 34
<212> PRT
<213> Homo sapien

<400> 137

Met Leu Leu Ile Arg Phe Tyr Leu Leu Phe Phe Ile His Arg Asp His
1 5 10 15

Lys Gln Ile Ala Asp Pro Gly Phe Ser Asn Trp Ser Ile Cys Leu Ile
20 25 30

Phe Pro

<210> 138
<211> 82
<212> PRT
<213> Homo sapien

<400> 138

Ser Leu Ser Val Ala Gln Ala Arg Val Gln Trp Arg Asp Pro Gly Ser
1 5 10 15

Leu Gln Pro Leu Pro Pro Gly Phe Lys Arg Phe Leu Ser Leu Ser Leu
20 25 30

Pro Ser Ser Ala Gly Tyr Arg Arg Ala Pro Pro Pro Cys Pro Ala Leu
35 40 45

Leu Tyr Phe Ala Val Glu Thr Gly Phe His His Val Gly Gln Ala Gly
50 55 60

Leu Glu Leu Leu Thr Ser Gly Asn Pro Ala Pro Pro Arg Pro Pro Lys
65 70 75 80

Val Leu

<210> 139
<211> 26
<212> PRT
<213> Homo sapien

<400> 139

Met Leu Asn Ser Phe His Val Phe Leu Asn Gln Leu Thr Asn Asn Phe
1 5 10 15

Glu Leu Val Ile Ser Ile Leu Gly Leu Ile
20 25

<210> 140
<211> 26
<212> PRT
<213> Homo sapien

<400> 140

Met Thr Ser Ile Pro Ser Ala Pro Gly Glu Lys Pro Gly Pro Arg Pro
1 5 10 15

Asp Pro Leu Lys Pro Asn His Ser Ser Phe
20 25

<210> 141
<211> 51
<212> PRT
<213> Homo sapien

<400> 141

Val Cys Gly Gly Ser Arg Gln Arg Gln Gly Leu Ala Pro Leu Ser Arg
1 5 10 15

Leu Glu Cys Phe Gly Val Met Thr Ala His Val Asn Leu Glu Phe Leu
20 25 30

Gly Ser Gly Asp Pro Pro Thr Ser Ala Ser Ala Leu Ala Glu Thr Thr
35 40 45

Gly Thr Arg
50

<210> 142
<211> 58
<212> PRT
<213> Homo sapien

<400> 142

Met Leu Gln Ala Arg Pro Pro Ala Ser Gly Lys Asn Gln Asn Thr Thr
1 5 10 15

Leu Lys Gly Gln Pro Ser Leu Gln Pro Ser Pro Cys Arg Glu Pro Ser

20 25 30

Leu Ala Leu Cys Cys Ser His Arg Ser Val Ser Gly Leu Ser Gln Val
35 40 45

Glu Gly Thr Cys Leu Thr Arg His Leu Cys
50 55

<210> 143
<211> 16
<212> PRT
<213> Homo sapien

<400> 143

Met Tyr Leu Arg Asp His Leu His Thr Ser Thr Ala Phe Val Cys Arg
1 5 10 15

<210> 144
<211> 84
<212> PRT
<213> Homo sapien

<400> 144

Met Arg Gln Ser Ala Thr Leu Arg Ser Ser Asp His Trp Glu Glu Arg
1 5 10 15

Glu Ser Leu Gln Leu Leu Gly Phe Arg Leu Gln Lys Phe Leu Ala Ala
20 25 30

Phe Ala His Trp Arg Gly Gly Glu Asp Lys Ser Ile Arg Asn Pro Met
35 40 45

Phe Pro Ser Ser Pro Thr Glu Arg Thr Lys Glu Val Phe Thr Arg Cys
50 55 60

Gly Thr Phe Leu Gln Leu Leu Asp Ala Asp Lys Pro Gln Ser Arg Leu
65 70 75 80

Phe Trp Leu Gln

<210> 145
<211> 88
<212> PRT
<213> Homo sapien

<400> 145

Met Ala Leu Glu Pro Gly Val Val Val Gln Val Leu Trp Arg Pro Ser
1   5   10   15

Tyr Ile Met Arg Leu Glu Ala Leu Arg Ile Ser Leu Ser His Gln Arg
            20   25   30

Ser Arg Leu Gln Trp Ala Arg Asp Trp Pro His Cys Ala Pro Ala Trp
        35   40   45

Val Thr Glu Pro Asn Val Val Ser Lys Lys Lys Lys Lys Lys Lys Lys
    50   55   60

Ala Ser Tyr Leu Pro Glu Val Ala Thr Pro Phe Leu Leu Ala Glu Ala
65   70   75   80

Gln Leu Gly Leu Thr Cys Pro Gly
                85

<210> 146
<211> 52
<212> PRT
<213> Homo sapien

<400> 146

Met Leu Leu Leu Gly Asn Met Thr Asn Pro Phe Glu Ala Thr Asn Phe
1   5   10   15

Met Ser Ser Phe Lys Ser Pro Ile Val Val Ile Phe Arg Lys Tyr Tyr
            20   25   30

Leu Thr Tyr Ser Met Ser Asn Ile Asn Leu Ile Lys Ser Leu Tyr Asn
        35   40   45

Ser Lys Lys Thr
    50

<210> 147
<211> 56
<212> PRT
<213> Homo sapien

<400> 147

Met Ser Ile Gly Val Ile Val Trp Thr Arg Gly Arg Val Pro Ile Val
1   5   10   15

Pro Pro Ser Glu Tyr Asp Gly Ser Cys Gly Thr Ala Arg Ser Ile Ala
20 25 30

Ala Cys Ser Arg Arg Arg Val Asn Val Arg Leu Gln Gly Phe Glu Pro
35 40 45

Ile His Phe Gln Leu Arg Cys Ile
50 55

<210> 148
<211> 92
<212> PRT
<213> Homo sapien

<400> 148

Met Ser Ala Leu Asn Pro Gly Gly Gln Arg Gly Val Tyr Glu Ala Arg
1 5 10 15

Val Pro Pro Thr Pro Thr Arg Gly Pro Lys Gly Ala Leu Pro Lys Lys
20 25 30

Lys Gln Gln Gln Gln Lys Cys Thr Asp Pro Ala Cys Thr Arg Leu Arg
35 40 45

His Ala Ser Leu Pro Ser Val Arg Leu Asp Pro Pro Pro Pro Ala Cys
50 55 60

Ile Lys Ser Gly Pro His Pro Pro Gly Arg Arg Ser Ile His His Met
65 70 75 80

Ala Pro Leu Glu His Asp Leu Glu Glu Gln Arg Leu
85 90

<210> 149
<211> 22
<212> PRT
<213> Homo sapien

<400> 149

Met Val Val Lys Asp His Leu Gly Ser Gln Gly Val Glu Gly Gly Gly
1 5 10 15

Ile Gln Phe His Arg Lys
20

<210> 150

<211> 254
<212> PRT
<213> Homo sapien

<400> 150

Met Glu Phe Pro Lys Met Leu Thr Arg Lys Ile Lys Leu Trp Asp Ile
1 5 10 15

Asn Ala His Ile Thr Cys Arg Leu Cys Ser Gly Tyr Leu Ile Asp Ala
20 25 30

Thr Thr Val Thr Glu Cys Leu His Thr Phe Cys Arg Ser Cys Leu Val
35 40 45

Lys Tyr Leu Glu Glu Asn Asn Thr Cys Pro Thr Cys Arg Ile Val Ile
50 55 60

His Gln Ser His Pro Leu Gln Tyr Ile Gly His Asp Arg Thr Met Gln
65 70 75 80

Asp Ile Val Tyr Lys Leu Val Pro Gly Leu Gln Glu Ala Glu Met Arg
85 90 95

Lys Gln Arg Glu Phe Tyr His Lys Leu Gly Met Glu Val Pro Gly Asp
100 105 110

Ile Lys Gly Glu Thr Cys Ser Ala Lys Gln His Leu Asp Ser His Arg
115 120 125

Asn Gly Glu Thr Lys Ala Asp Asp Ser Ser Asn Lys Glu Ala Ala Glu
130 135 140

Glu Lys Pro Glu Glu Asp Asn Asp Tyr His Arg Ser Asp Glu Gln Val
145 150 155 160

Ser Ile Cys Leu Glu Cys Asn Ser Ser Lys Leu Arg Gly Leu Lys Arg
165 170 175

Lys Trp Ile Arg Cys Ser Ala Gln Ala Thr Val Leu His Leu Lys Lys
180 185 190

Phe Ile Ala Lys Lys Leu Asn Leu Ser Ser Phe Asn Glu Leu Asp Ile
195 200 205

Leu Cys Asn Glu Glu Ile Leu Gly Lys Asp His Thr Leu Lys Phe Val

210 215 220

Val Val Thr Arg Trp Arg Phe Lys Lys Ala Pro Leu Leu Leu His Tyr
225 230 235 240

Arg Pro Lys Met Asp Leu Leu Arg Pro Lys Met Asp Leu Leu
245 250

<210> 151
<211> 40
<212> PRT
<213> Homo sapien

<400> 151

Met Gly Thr Arg Tyr Tyr Ile Leu Glu Phe Val Leu Arg Arg His Lys
1 5 10 15

Leu Asn Ser Arg Ser Leu Cys Pro Lys Phe His Arg Leu Lys Lys Arg
20 25 30

Ser Ser Asn Tyr Arg Ser Gly Tyr
35 40

<210> 152
<211> 42
<212> PRT
<213> Homo sapien

<400> 152

Met Glu Asn Ser Gln Glu Met Asn Glu Lys Arg Leu Cys Glu Ser Tyr
1 5 10 15

Ala Thr Val Tyr Ile Thr Ser Cys Glu Ala Ile Arg Leu Lys Thr Arg
20 25 30

Ala Asn Leu Lys Thr Lys Leu Phe Ser Cys
35 40

<210> 153
<211> 51
<212> PRT
<213> Homo sapien

<400> 153

Met Leu Leu Ser Tyr Ile Ser Gly Arg Phe Leu Ser Thr Arg Lys Glu
1 5 10 15

Asn Thr Gly Leu Ala Lys Gln Gly Pro Leu Phe Gly Ile Ile Phe Val
20 25 30

Pro Asn Lys Gln Ser Arg Gly Trp Val Cys Trp Leu Val Lys Glu Leu
35 40 45

Leu Arg Phe
50

<210> 154
<211> 63
<212> PRT
<213> Homo sapien

<400> 154

Met Leu Glu Pro Ala Ala Ser Met Ile Gly Met Pro Gly Gln Val Gly
1 5 10 15

Ser Arg Gly Gly Cys Ser Asp Arg Arg Val His Ser Ser Tyr Asn Arg
20 25 30

Gly Val Leu Asp Phe Ile Leu Gln Ser Glu Leu Ser Thr Phe Ala Phe
35 40 45

Trp Arg Thr Gln Val Thr Ala His Leu Pro Phe Leu Leu Glu Pro
50 55 60

<210> 155
<211> 50
<212> PRT
<213> Homo sapien

<400> 155

Met Lys Pro Lys Lys Lys Lys Lys Arg Gln Lys Lys Arg Val Leu Trp
1 5 10 15

Gly Asn Pro Gly Gly Leu Arg Met Cys Ser Leu Val Cys Arg Thr Ile
20 25 30

Val Val Pro Val Pro Asn Phe Pro Pro Tyr Ser Ser Val Asp Asp Lys
35 40 45

Arg Gly
50

<210> 156
<211> 35
<212> PRT
<213> Homo sapien

<400> 156

Met Phe Tyr Leu Gly Phe Arg Val Asn Lys Lys Lys Lys Thr Cys Val
1 5 10 15

Leu Ser Phe Cys Asp Arg Thr Glu His Ile Thr Arg Arg Lys Arg Gly
20 25 30

Gly Arg Lys
35

<210> 157
<211> 73
<212> PRT
<213> Homo sapien

<400> 157

Met Gly Arg Cys Ser Leu Phe Thr Pro Ala Ala Ile Gly Glu Arg Gly
1 5 10 15

Ile Gln Leu Ile Ser Tyr Leu Tyr Arg Met Asp Tyr Leu Cys Lys Asn
20 25 30

Lys Asn Leu Gln Thr Lys Asp Ile Val Glu Leu His Tyr Pro Pro Ser
35 40 45

Gln Asp Glu Ser Thr Asp Met Gln His His Asp His Glu Gln Met Val
50 55 60

Pro Leu Gly Met Pro Met Val Gly His
65 70

<210> 158
<211> 82
<212> PRT
<213> Homo sapien

<400> 158

Met Tyr Leu Ser Val Cys Val Cys Val Cys Val Cys Tyr Gly Gly Arg
1 5 10 15

Gly Gly Phe Phe Lys Ile Ser Val Val Cys Gly Phe Phe Phe His Thr
20 25 30

Leu Val Pro Thr Ile Ala Cys Pro Gly Thr Thr Ala Trp His Tyr Arg
35 40 45

Met Leu Gly Ser Ser Gly Glu Gly Ser Glu Ala His His Cys Pro Leu
50 55 60

Phe Trp Arg Phe Leu Phe Leu His Lys Val Val His Pro Ile Gln Ile
65 70 75 80

Ala Asp

<210> 159
<211> 82
<212> PRT
<213> Homo sapien

<400> 159

Met Leu Asn Thr Cys Arg Val Ile Leu Val Val Phe Ser Gln Pro Phe
1 5 10 15

Ile Lys Phe Leu Val Thr Ser Val Met Met Thr Phe His Thr Pro Ile
20 25 30

Thr Ser Lys Ala Phe Leu His Leu Ala Asp Pro Ser Tyr Gly Pro Ala
35 40 45

Val Ser His Ala Val Thr Thr Ser Gly Thr Asp Leu Thr Ala Leu Arg
50 55 60

Ala Ser Ser Ser Leu Ala Gly Arg Thr Ser Ala Ala Ser Ser Ile Thr
65 70 75 80

Lys Gly

<210> 160
<211> 200
<212> PRT
<213> Homo sapien

<400> 160

Met Arg Arg Lys Arg Lys Thr Arg Leu Ser Val Arg Pro Gly Ser Glu
1 5 10 15

Leu Ser Lys Leu Pro Arg Leu Ala Leu Asn Gln Asn His Phe Ala Ser
20 25 30

Gln Pro Arg Pro Leu Gly Tyr Thr Ala Leu Asn Gly Pro Ala Asn Ala
35 40 45

Gly His Ser Ile Ser Leu Val Leu Glu Thr Arg Glu Leu Lys Gln Ser
50 55 60

Ile Pro Leu Ser Asn Lys Ile Met Asp Ser Ala Lys Lys Lys Gln Lys
65 70 75 80

Lys Lys Lys Gly Cys Gly Gly Thr Pro Gly Ala Ile Arg Gly Pro Gly
85 90 95

Cys Glu Leu Val Ser Arg Ser Ile His Ser Asp Thr His Thr Ser Arg
100 105 110

Lys Lys Lys Glu Glu Asn Thr Ser Glu Lys Arg Lys Asn Thr Thr Arg
115 120 125

Arg Lys Lys Lys Pro Glu Lys Ala Thr Arg Lys Gln Arg Glu Asn Lys
130 135 140

Arg Ala Arg Gly Lys Arg Asp Ala Arg Lys Lys Lys Gln Glu Pro Gln
145 150 155 160

Ala Glu Thr Glu Thr Ser Lys Gly Thr Gln Arg Arg Thr Thr Lys Arg
165 170 175

Ser Gln Glu Gln Thr Lys Ala Arg His Lys Ala Asp Asp Glu Arg Gly
180 185 190

Thr Arg Lys Glu Arg Lys Arg Glu
195 200

<210> 161
<211> 38
<212> PRT
<213> Homo sapien

<400> 161

Met Asp Ala Trp Ser Arg Arg Gly Thr Glu Ser Cys Tyr Phe Ser Leu
1 5 10 15

Arg Pro Tyr Leu Ala Ala Phe Ile Asn Ala Ser Glu Leu Tyr Val Ile
20 25 30

Ile Ile Trp Ile Tyr Thr
35

<210> 162
<211> 66
<212> PRT
<213> Homo sapien

<400> 162

Met Asp Ala Gln Trp Ser Gly Arg Ser Asp Val Trp Ser Ser Glu Val
1 5 10 15

Glu Lys His Glu Ser Lys Asp Gln His Leu Gly Val Leu Leu Leu Cys
20 25 30

Leu Val Asn Arg Gly Leu Arg Ala Val Phe His Leu Val Pro Phe Ser
35 40 45

Glu Asp Gln Ile Pro Arg Leu Gln Ser Met Gln Gly Leu His Arg Trp
50 55 60

Leu Leu
65

<210> 163
<211> 76
<212> PRT
<213> Homo sapien

<400> 163

Met Gly Glu Leu Gly Arg Glu Thr Lys Phe His Pro Gly Pro Leu Trp
1 5 10 15

Pro Arg Val Pro Gln Ala Phe Phe Phe Phe Val Phe Phe Phe Phe Arg
20 25 30

Leu Leu Met Asp Leu Gln Arg Leu Glu Gln Pro Phe Arg Gln Thr Gln
35 40 45

Val Thr Ser Ile Glu Ser Leu Leu Asn Leu Ser Glu Ile Tyr Met Leu
50 55 60

Glu Leu Gln Val Asn Ser Pro Val Asn Thr Gln Ala
65 70 75

<210> 164
<211> 69
<212> PRT
<213> Homo sapien

<400> 164

Met His Val Pro Met Arg Glu Ser Met His Val Cys Ala Tyr Glu His
1 5 10 15

Lys Leu Leu Cys Trp Arg Gly Ser Trp Glu Arg Arg Gly Glu His Ala
20 25 30

Leu Leu Val Ile His Ile His Ser Tyr Val Cys Thr His Asn Ile His
35 40 45

Pro Glu Pro Val Ser Gln Ile Asp Gly Ser Lys Ser Leu Ser Tyr Arg
50 55 60

Arg Pro Asp Pro Thr
65

<210> 165
<211> 53
<212> PRT
<213> Homo sapien

<400> 165

Met Leu Pro Phe Ser Gly Leu Val Tyr Thr Leu Phe Phe Val Phe Phe
1 5 10 15

Phe Val Arg Gln Asp Leu Ala Leu Ser Ala Arg Leu Glu Cys Ser Gly
20 25 30

Thr Gly Met Ile His Cys Arg Thr Pro Gly Leu Lys Arg Phe Thr Cys
35 40 45

Leu Lys Pro Leu Met
50

<210> 166
<211> 86
<212> PRT
<213> Homo sapien

<400> 166

Glu Thr Gly Ser Cys Ser Val Cys Gln Ala Gly Val Gln Trp His Arg
1 5 10 15

Tyr Asp Ser Leu Gln Asn Ser Trp Ala Gln Glu Ile His Leu Pro Ala
20 25 30

Ala Ser His Val Ala Gly Asp His Ser Ala Tyr Gly His Thr Trp Cys
35 40 45

Leu Gln Pro His Leu Ala Asn Phe Leu Phe Phe Phe Asn Gly Asn Lys
50 55 60

Val Ser Leu Cys Cys Pro Val Trp Ser Ala Thr Pro Glu Ile Gln Arg
65 70 75 80

Ser Ser His Leu Gly Ile
85

<210> 167
<211> 52
<212> PRT
<213> Homo sapien

<400> 167

Met Glu Arg His Gly Glu Ile Phe Leu Pro Thr Leu Asn Tyr Ser Asn
1 5 10 15

Tyr Ser Lys Thr Ser Asn Leu Lys Thr Asn Arg Arg Ser Pro Thr Gly
20 25 30

Leu Lys Arg Arg Met Arg Asp Lys Glu Lys Ser Val Trp Leu Pro Leu
35 40 45

Leu Ser Thr Asp
50